Introducción analítica a las geometrías

Javier Bracho

August 9, 2005

Abstract

A Irene, Felipe y Adrián.

A mis maestros:
Victor Neumann Lara,
Sevín Recillas Pishmish.

Prefacio

Este libro es un texto para el primer nivel universitario. No presupone "cierta madurez matemática" sino que se empeña en crearla, y sólo requiere del estudiante conocimientos mínimos del lenguaje de teoría de conjuntos, además de su esfuerzo. Fue concebido y experimentado en cursos de geometría analítica de primer año de las licenciaturas en matemáticas, física, actuaría y ciencias de la computación; pero puede extenderse a otras áreas y emplearse en otros niveles.

A principios del siglo XXI, la geometría —como área de las matemáticas— evade las definiciones pues sus límites son difusos y sus ramificaciones numerosas. Es, quizá, más que un cuerpo bien definido de conocimiento: una sensibilidad al practicar el pensamiento abstracto. El título de "geómetra" hoy se lo disputan matemáticos de muy diversas áreas; por dar ejemplos, los que hacen investigación en geometría algebraica, geometría diferencial o geometría discreta se llaman a sí mismos "geómetras", con orgullo, a secas y con razón, pero a veces también con soberbio aire excluyente. Además, a cada rato aparecen nuevas sectas de "geómetras" que abren y bautizan áreas como "geometría computacional", "geometría simpléctica" o "geometría no conmutativa", por no hablar del uso del término como calificativo: "topología geométrica", "convexidad geométrica", "teoría geométrica de grupos", "física geométrica", etc. Quiero dar a entender con esta breve enumeración —que sólo a unos pocos lectores dice algo más que la rimbombancia misma de las palabras— que la Geometría, ahora sí, con mayúsculas, está vivita y coleando: en plena expansión como el resto de la ciencia. Sea lo que fuere, área o sensibilidad, es un motor pujante, activo y prolífico en el imponente desarrollo actual de las matemáticas. Es, en fin, una parte medular de la matemática contemporánea, pero a su vez la rebasa como sensibilidad o por sus aplicaciones, la desborda en sus manifestaciones naturales o artísticas; es una gema de nuestra cultura.

Sin embargo, y en contraste preocupante con este panorama, la educación de la geometría en el nivel universitario básico ha quedado estancada y ha sido relegada casi hasta la vacuidad. En muchas universidades, el curso clásico de "geometría analítica" ha desaparecido o se ha integrado a los cursos de cálculo como servicio, pues las cónicas —argumentan— son de nivel bachillerato; y en aquellas donde aún se da es con programas obsoletos que parecen ignorar que el siglo XIX aconteció. Los cursos de "geometría sintética" o "a la griega", a veces llamados con ironía involuntaria "geometría moderna" —pues tratan de "geometría griega" hecha despues de ellos— , también se sienten anquilosados y, en el mejor de los casos, se usan como materias optativas de divertimento histórico, didáctico o formativo, desligados por completo del resto del currículo científico. No obstante, se siguen formando geómetras pues la sensibilidad geométrica es irrefrenable. Los estudiantes que la tienen —don, vocación, inquietud o visión del mundo— aprenden por ósmosis y desarrollan su intuición en otros cursos básicos o intermedios, como el de álgebra lineal o

topología, y sólo hasta niveles avanzados de licenciatura o maestría vuelven a encontrar "geometrías", pero ya con apellidos de abolengo como "algebraica", "riemanniana" o "diferencial", o en boga como "cuántica" y "computacional", ya muy especializadas.

Este contraste entre la vitalidad de la geometría en el desarrollo de las matemáticas —su enorme importancia en ellas— por un lado, y por el otro, su magro valor curricular en la formación básica de científicos y profesionistas —su ausencia en la cultura ciudadana mínima—, obliga a una reflexión seria, sobre todo de parte de los que nos consideramos geómetras.

Para arañar la superficie de esta discusión, tengo que recurrir al balance entre las dos grandes tendencias, fuerzas o procesos que desarrollan la ciencia (cuya frontera es difícil dibujar.) Se tiene por un lado la fuerza creativa o de investigación que la hace avanzar y conquistar nuevo conocimiento. Pero además, hay un proceso de digestión, destilación o decantación en el cual nos vamos quedando con lo fundamental de las grandes teorías al reprocesar las ideas básicas, limpiar los argumentos centrales o aclarar y simplificar los razonamientos y construcciones; de tal manera que cada generación tiene acceso a ese conocimiento a una edad más temprana y con más profundidad que sus predecesoras.

En el siglo XX el proceso de digestión de las matemáticas se centró en sus fundamentos y formalización. Se creó un lenguaje, un estilo y una notación basados en la Teoría de Conjuntos (iniciada por Georg Cantor hacia finales del XIX) que permeó y cambió a todas las matemáticas; desde cómo se presentan a los párvulos hasta cómo se escriben sus artículos de investigación, pasando, por supuesto, por sus libros de texto y los currículos de todos los niveles de enseñanza. Pecando de simplismo pero, y valga la redundancia, en aras de simplificar mi argumento, permítaseme ubicar este monumental cambio en el ámbito de lo *simbólico*: "la otra sensibilidad" para enfrentar las matemáticas. Me aventuro a afirmar entonces que la digestión de la geometría quedó rezagada porque la energía intelectual de un siglo no puede desperdigarse demasiado. Pero la marea, como el siglo, está cambiando.

En los últimos 20 o 30 años del siglo XX, se dio un renacimiento de la geometría del siglo XIX en el nivel de investigación que se está digiriendo rápidamente y permeando en la enseñanza. La geometría hiperbólica renació ligada a la topología en dimensiones bajas; la geometría proyectiva revivió como fundamento para desarrollar la visualización por computadora —y ésta a su vez revaloró la parte algorítmica de la geometría (las construcciones con regla y compás que apasionaban a los griegos)—; los grupos geométricos encontraron excitantes aplicaciones en la física, etc. Y en este entorno, han aparecido nuevos libros de texto que dan a la geometría un enfoque más "kleiniano". Lo llamo así porque al concluir el siglo XIX —en el que se convulsionó la geometría al descubrirse sus versiones no euclidianas— la "Geometría" había dejado de ser una unidad, se había "desvalagado" en múltiples direcciones al parecer divergentes y Felix Klein, en su famosa ponencia del *Programa de Erlangen* y con profundidad visionaria, trató de resumir esta revolución diciendo, a muy

grandes rasgos, que “la geometría es el estudio de espacios junto con sus grupos de transformaciones”. Casi un siglo después, empezó a plasmarse esta filosofía en textos de nivel medio universitario.

De estos nuevos libros de texto, tuvieron gran influencia sobre mí, y por lo tanto en el presente libro, el de E. Rees [12] y el de P. J. Ryan [14]. Sin embargo, presuponen ya una cierta madurez matemática de parte del estudiante; la filosofía de este libro (enfatizar los grupos de transformaciones) es muy parecida pero llevada a un nivel más elemental, introductorio. Se desarrolló en los cursos de Geometría Analítica que periódicamente he dado en la Facultad de Ciencias de la UNAM durante 18 años. Con un programa oficial que bien podría datar del siglo XVIII (la geometría analítica como se entendía entonces), este curso ha ido evolucionando hasta su presente forma. La condición inicial de tener que cubrir un programa basado en el estudio de las cónicas, más que estorbar, acabó dándole una gran consistencia. Pues los problemas que plantean motivan el desarrollo teórico que a su vez, con más herramientas geométricas en mano, permiten entenderlas más profundamente y plantear nuevos problemas que a su vez... así que quedaron tratadas en cuatro capítulos (2, 4, 7 y 9).

Es frecuente pensar (yo lo hice por demasiado tiempo) que la geometría analítica representa un rompimiento con la geometría griega clásica, presentarla como algo que la supera; o bien, creer que las geometrías proyectiva e hiperbólica rompen con la euclidiana y que hay que tomar partido por alguna. No. Son todas ellas parte de lo mismo: La Geometría. Y para hacer énfasis en ello, una preocupación que entreteje al capítulo 1 es reconstruir la axiomática euclidiana a partir de la de los números reales que es la de la geometría analítica, y aclarar formalmente su relación. Parece superfluo, pero permite ejercitar las operaciones vectoriales y la idea de “demostración” en casos sencillos e intuitivamente claros, va construyendo la llamada “madurez matemática” y en los ejercicios, tanto teóricos como númericos, debe dar al estudiante seguridad y fe en su intuición geométrica; el riesgo de parecer trivial quizá se conjure con la idea profunda de que el lenguaje y las técnicas vectoriales que se van estableciendo en el plano funcionarán también para las siguientes dimensiones. Más adelante (capítulos 6 y 8, respectivamente), como se construyen los planos proyectivo e hiperbólico, ya no hay necesidad de axiomatizar; vamos sobre terreno firme y los axiomas quedan como observaciones o teoremas con notas históricas.

Si insisto en que La Geometría es una y única, ¿por qué pluralizarla en el título del libro? Por dos razones. Así como en español tenemos la afortunada dualidad “la matemática” y “las matemáticas” con significados sutilmente diferentes, siento que lo mismo se aplica a “la geometría”. “Las geometrías” tendría entonces una acepción más modesta, con una implícita diversidad de enfoques y motivaciones posibles; pretendo, quizá pecando de ambicioso e iluso, que aquí se den los fundamentos modernos para adentrarse en ellas y en especial en las que están vivas en este cambio de siglo. Pero además, el plural tiene la ventaja de que remite inevitablemente a las geometrías no euclidianas, cuyos rudimentos básicos, creo yo, deberían ser parte de la formación temprana en

matemáticas, e integrarse a la cultura mínima común sobre ellas. Sin embargo, por "las geometrías" no quiero entender sólo a las tres clásicas sino algo más cercano a la enumeración del principio. Al final de varios capítulos se incluyen secciones marcadas con un asterisco que indica que no es esencial para el desarrollo ulterior del libro; son pequeñas introducciones o disertaciones hacia áreas o temas aledaños, hacia "otras geometrías". Están ahí para incitar la curiosidad de los estudiantes o para dar más material a los maestros que las sientan afines. Por ejemplo, al final del capítulo 1, después de haber insistido en la maravillosa idea de Descartes de asociarle numeritos a los puntos que forman un plano y de haber hecho escencialmente lo mismo con las rectas mediante los coeficientes de sus ecuaciones, está a la mano preguntarse ¿geométricamente, qué forman las rectas del plano? Surge la banda de Moebius, una bonita motivación para entrar a la topología: gloria indiscutible de La Geometría del siglo XX.

En geometría, y sobre todo en el nivel elemental pero con una persistencia sorprendente hasta niveles más abstractos, los dibujos son parte fundamental del discurso matemático, del razonamiento, del proceso de entender y descubrir; quizá lo que tenga más que ver con eso de la "sensibilidad geométrica" que tanto he cacareado; son básicos para "visualizar" —yo no puedo prescindir de ellos en el salón de clase—. Pero sabemos bien que no son más que muletas para acceder a las verdaderas matemáticas. Una figura al margen o en el pizarrón es —en el mejor de los casos y de no ser un bosquejo esquemático de algo mucho más complicado— una instancia muy particular y aproximada de un razonamiento, de una definición o, digamos, de un cierto teorema. Las computadoras permiten liberar un poco esta particularización o congelamiento de los dibujos. Podemos tener *figuras animadas* que den la idea de un proceso continuo o que, al menos, hagan variar las instancias particulares donde se cumple el teorema. Pero también podemos tener *figuras interactivas* donde el "usuario" mueve los parámetros u objetos variables a los que alude la figura al mismo tiempo que ve en pantalla los efectos de sus actos, y entonces (en principio) puede pasar por todas las instancias a las que se refiere el teorema. Ambas muletas se acercan un poco más al teorema en sí; al menos amplían los ejemplos donde se cumple y ciertamente proveen una nueva intuición que debe ser bienvenida y explotada. Adjuntamos un disco compacto con una selección de figuras del libro a las que se dotó de algún don dinámico. Con una computadora se puede jugar con ellas —me sorprendió, en algunas, lo lejos que mi intuición andaba de la realidad— y quizá coadyuven a generar nuevas maneras de relacionarse con la geometría y las matemáticas. Fueron creadas con el programa Cinderella© y son, en su gran mayoría, construcciones clásicas con regla y compás; sus códigos, programas o algoritmos también vienen en el CD y pueden entonces servir como ejemplos didácticos de este recién resucitado arte que, como parte fundamental de la geometría, nació con los griegos.

Es un hecho que el enfoque (históricamente) original con el que se resuelven los problemas no es siempre el más directo, conciso, claro o elegante. Lo mismo sucede con las grandes teorías; su enseñanza siguiendo estrictamente su línea de desarrollo histórico no es necesariamente la más profunda o eficiente. Estoy

convencido de que la cultura geométrica general, y muy en especial la de los científicos y matemáticos, debe actualizarse con urgencia para tener un rezago de por lo más un siglo, no de cuatro, y esto requiere buscar caminos e itinerarios alternativos que lleven de manera más simple y expedita a la geometría que se cocina hoy día. Este libro es fruto de mi esfuerzo para contribuir a ello.

Prólogo para el maestro

Aunque este libro esté pensado para un curso de un año, creo que en dos semestres es casi imposible cubrir todo su material. Yo nunca lo he logrado. Basicamente por las "secciones asterisco" de las que hablé antes, esos divertimentos laterales y aledaños sobre los cuales trabajo con distinta (o nula) profundidad dependiendo de los tiempos del curso, el grupo o mis intereses del momento. Sin embargo, siento que es importante que estén ahí. Si el estudiante decide brincarse olímpicamente alguno de ellos porque no es material del examen que le apremia, aprenderá que los libros de matemáticas pueden leerse a saltos, a veces grandes para adelante y cuando es necesario para atrás. Y si los lee por su cuenta, que sí creo posible que se haga con todas, quizá estimule su curiosidad o creatividad; al menos ampliará su cultura. En fin, más vale que sobre y no que falte, que se abran panoramas y no dar la impresión de algo ya hecho, estático y sin puertas o futuros posibles.

Sin embargo, sí creo que el material básico puede cubrirse en un año. Un semestre para el plano euclidiano, sus transformaciones y, con éstas últimas como herramienta, la clasificación de las cónicas; y el segundo para la esfera y el espacio, los planos proyectivo e hiperbólico y algo más de cónicas como apoyo, herramienta (en la construcción del plano hiperbólico) y motivación. Por supuesto, los énfasis o la profundidad con que se vean algunas partes no esenciales depende de los gustos y necesidades de cada maestro.

Me imagino también que los capítulos 3 (a "salto de mata"), 6 (con referencias a parte del 5) y 8 (con las pocas definiciónes que usa del 7) pueden ser el material de un curso semestral intermedio de geometrías no euclidianas para estudiantes que sepan geometría analítica clásica, algo de álgebra lineal y algo de teoría de grupos; el resto del libro será repaso o "asterisco".

De álgebra líneal se da necesariamente una introducción (capítulos 1 y principio del 5), pero es lo mínimo para hacer geometría en dimensiones 2 y 3; servirá de motivación y experiencia para un curso general. Los grupos en abstracto apenas se mencionan tangencialmente, de hecho ni siquiera se definen formalmente, pues el interés está únicamente en los grupos geométricos de transformaciones. Darán al estudiante una multitud de ejemplos e intuición para enfrentarse, en su momento, al álgebra moderna, en especial las secciones asterisco sobre grupos geométricos finitos y discretos. Puede entonces servir como texto de apoyo para materias más algebráicas.

La naturalidad en el orden en que se presentan las ideas matemáticas es, muchas veces, subjetiva. Cada maestro o cada autor siente esto a su manera; "las matemáticas no son simplemente conexas". Por ejemplo, para alguien que

sabe, parece una locura hablar de transformaciones ortogonales antes de las lineales como aquí se hace; la razón es que las isometrías están más cerca de la intuición de alguien que no sabe de ninguna de ellas y al seguir el razonamiento desde esta óptica se obtiene, naturalmente creo yo, la noción de ortogonalidad y, con ésta, la linealidad como resultado: ello se convierte en motivación para definirlas. Por supuesto, el maestro que siga el orden clásico (más natural si de antemano se sabe de linealidad), o el de su gusto, y use a la vez este texto, enriquecerá la visión del estudiante.

En algunos otros lados del libro pasa lo mismo, hay alteraciones al orden que, en décadas pasadas, se nos enseñó como natural. No hay espacio para dar mis razones particulares. Pero en general, puedo decir que se deben a tres razones básicas: dar al texto consistencia interna y hacerlo, en la medida de lo posible, autocontenido; tratar de seguir líneas de razonamiento con motivaciones generales precisas y buscar dar ejemplos relevantes después de las definiciones para familiarizarse con ellas (por ejemplo, las coordenadas baricéntricas se introducen tan temprano para presentar el Teorema del Baricentro y así jugar con combinaciones lineales.)

Prólogo para el estudiante

Aprender matemáticas es un acto de soledad. Escuchar a un maestro o leer un libro ayudan y quizá explicar a alguien, o discutir, sirva más; pero a fin de cuentas requiere un proceso íntimo de creación y recreación. Por eso, leer matemáticas tiene sus propios ritmos o mañas que cada quien desarrolla. Una buena es la del lápiz y papel a la mano, dándoles su tiempo; otra, la de leer tramos rápido y por ensima para luego regresar, si acaso, a los detalles. Algo importante en un libro de texto como éste es entablar, al menos, breves escaramuzas mentales con los ejercicios. Hay que notar que algunos de ellos tienen un asterisco, y eso es porque son más dificiles que los otros; vale la pena preguntarse por qué. Y en general, ésa es la actitud correcta ante un texto de matemáticas: siempre hay que cuestoinarlo tratando de mejorarlo o corregirlo; no necesariamente su ritmo o hilo de razonamiento es el natural de uno, y encontrar esa "voz" es aprender. Un texto matemático, por decirlo así, nos presta su voz por un rato, pero aprender, o entender, es poder referirnos a las matemáticas a las que alude con nuestra propia voz, ser capaces de verlas con nuestros propios ojos.

Agradecimientos

Quiero agradecer en primer lugar a los estudiantes que llevaron conmigo el curso "Geometría Analítica" en alguna de sus tantas versiones preliminares. No era fácil, pues no sabía bien cómo llegar a donde iba: estaba aprendiendo junto con ellos y ensayando; a veces nos confundíamos todos pero siempre nos divertimos. Sus preguntas, comentarios o bien las simples caras de extravio fueron centrando el rumbo. Siento que les debía este libro.

En la misma medida, o más, agradezco el apoyo creativo de los ayudantes de esos cursos; no los enlisto por el temor a saltarme alguno. Tampoco era nada fácil pues había que responsabilizarse y el material era volátil. Su trabajo, más en corto conmigo, fué fundamental para asentar las ideas y la concepción. Muy en especial, debo agradecer a Ricardo Strausz que escribió la primera versión de los primeros capítulos y, sólo así, me obligo a hacer lo propio.

Ya con el manuscrito, transformarlo en libro fue trabajo de equipo. Agradezco a Gabriela Alfaro y Claudia ¿¿??, su ayuda con los archivos, índices y correcciones. A Abdiel Macias su meticulosa corrección de estilo. A Emiliano Mora por su trabajo en las figuras interactivas. A Guadalupe ¿¿?? y ¿¿?? por sus figuras; dan tal personalidad visual al libro que, desde ya, hacen que me trascienda. Por último, agradezco profundamente a mis editores María del Carmen Farías y Axel Retif por su intenso apoyo al proyecto y por la excelencia amorosa de su trabajo.

Contenido

Capítulo 1

El plano euclidiano

1.1 La geometría griega

En el principio, la geometría era una colección de reglas de uso común para medir y construir casas y ciudades. No fue hasta el año 300 a.C. que Euclides de Alejandría, en sus *Elementos*, ordenó y escribió todo ese saber, imprimiéndole el sello de rigor lógico que caracteriza y distingue a las matemáticas. Se dio cuenta de que todo razonamiento riguroso (o *demostración*) debe basarse en ciertos principios previamente establecidos ya sea, a su vez, por otra demostración o bien por convención. Pero a final de cuentas, este método conduce a la necesidad ineludible de convenir en que ciertos principios básicos (*postulados* o *axiomas*) son válidos sin necesidad de demostrarlos, que están dados y son incontrovertibles para poder construir sobre ellos el resto de la teoría. Lo que hoy se conoce como Geometría Euclidiana, y hasta hace dos siglos simplemente como Geometría, está basada en los cinco postulados de Euclides:

I Por cualesquiera dos puntos, se puede trazar el segmento de recta que los une.

II Dados un punto y una distancia, se puede trazar el círculo con centro en el punto y cuyo radio es la distancia.

III Un segmento de recta se puede extender en ambas direcciones indefinidamente.

IV Todos los ángulos rectos son iguales.

V Dadas dos rectas y una tercera que las corta, si los ángulos internos de un lado suman menos de dos ángulos rectos, entonces las dos rectas se cortan y lo hacen de ese lado.

Obsérvese que en estos postulados se describe el comportamiento y la relación entre ciertos elementos básicos de la geometría, como son "punto", "trazar", "segmento", "distancia", etc. De alguna manera, se le da la vuelta a su definición haciendo uso de

la noción intuitiva que se tiene de ellos y haciendo explícitas ciertas relaciones básicas que deben cumplir.

De estos postulados o axiomas, el quinto es el más famoso pues se creía que podría deducirse de los otros. Es más sofisticado, y entonces se pensó que debía ser un *teorema*, es decir, ser demostrable. De hecho, un problema muy famoso fue ése: demostrar el quinto postulado usando únicamente los otros cuatro. Y muchísimas generaciones de matemáticos lo atacaron sin éxito. No es sino hasta el siglo XIX, y para sorpresa de todos, que se demuestra que no se puede demostrar; que efectivamente hay que suponerlo como axioma, pues negaciones de él dan origen a "nuevas geometrías", tan lógicas y bien fundamentadas como la euclidiana. Pero esta historia se verá más adelante (en los capítulos 6 y 8) pues por el momento nos interesa introducir la geometría analítica.

La publicación de la "*Géométrie*" de Descartes marca una revolución en las matemáticas. La introducción del álgebra a la solución de problemas de índole geométrico desarrolló una nueva intuición y, con ésta, un nuevo entender de la naturaleza de las "*lignes courves*".

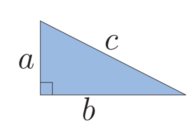
Para comprender lo que hizo René Descartes (1596–1650), se debe tener más familiaridad con el quinto postulado. Además del enunciado original, hay otras dos versiones que son equivalentes:

V.a (**El Quinto**) Dada una línea recta y un punto fuera de ella, existe una única recta que pasa por el punto y que es paralela a la línea.

V.b Los ángulos interiores de un triángulo suman dos ángulos rectos.

De las tres versiones que hemos dado del quinto postulado de Euclides, usaremos a lo largo de este libro a la **V.a**, a la cual nos referiremos simplemente como "*El Quinto*".

Antes de entrar de lleno a la Geometría Analítica demostremos, a manera de homenaje a los griegos y con sus métodos, uno de sus teoremas más famosos e importantes.

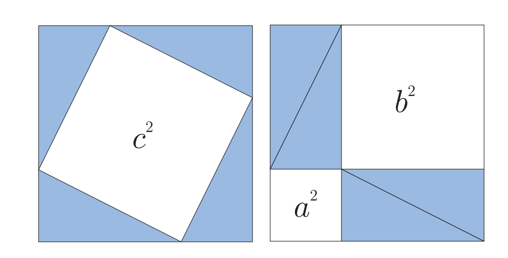


Teorema 1.1 (de Pitágoras) *Dado un triángulo rectángulo, si los lados que se encuentran en un ángulo recto (llamados* catetos*) miden* a *y* b*, y el tercero (la* hipotenusa*) mide* c*, entonces*

$$a^2 + b^2 = c^2.$$

Demostración. Considérese un cuadrado de lado $a + b$, y colóquense en él cuatro copias del triángulo de dos maneras diferentes como en la figura. Las areas que quedan fuera son iguales. □

Aunque la demostración anterior (hay otras) parece no usar nada, queda implícito el uso del quinto postulado; pues, por ejemplo, en el primer cuadrado, que el cuadradito interno de lado c tenga ángulo recto usa su versión **V.b**. Además, hace uso de nuevos conceptos como son el área y cómo calcularla; en fin, hace uso de nuestra intuición geométrica, que debemos creer y fomentar. Pero vayamos a lo nuestro.



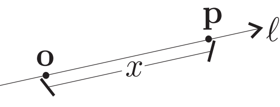
***EJERCICIO 1.1** Demuestra la equivalencia de **V**, **V.a** y **V.b**. Aunque no muy formalmente (como en nuestra demostración del Teorema de Pitágoras), convencerse con dibujos de que tienen todo que ver entre ellos.

EJERCICIO 1.2 Demuestra el Teorema de Pitágoras ajustando (sin que se traslapen) cuatro copias del triángulo en un cuadrado de lado c y usando que $(b-a)^2 = b^2 - 2ab + a^2$ (estamos suponiendo que $a < b$, como en la figura anterior).

1.2 Puntos y parejas de números

Para reinterpretar el razonamiento que hizo Descartes, supongamos por un momento que conocemos bien el *plano euclidiano* definido por los cinco axiomas de los *Elementos*; es el conjunto de puntos que se extiende indefinidamente a semejanza de un pizarrón, un papel, un piso o una pared y viene acompañado de nociones como rectas y distancia, entre otras; y en el cual se pueden demostrar teoremas como el de Pitágoras. Denotaremos por $\mathbb{E}^2$ a este plano; en este caso el exponente 2 se refiere a la dimensión y no es exponenciación (no es "$\mathbb{E}$ al cuadrado", sino que debe leerse "$\mathbb{E}$ dos"). Da la idea de que puede haber espacios euclidianos de cualquier dimensión $\mathbb{E}^n$ (léase "$\mathbb{E}$ ene"), aunque sería complicado definirlos axiomáticamente. De hecho los definiremos, pero de una manera más simple, que es con el uso de coordenadas: la idea genial de Descartes. Podríamos ahora, apoyados en el lenguaje moderno de los conjuntos,[1] recrear su razonamiento como sigue.

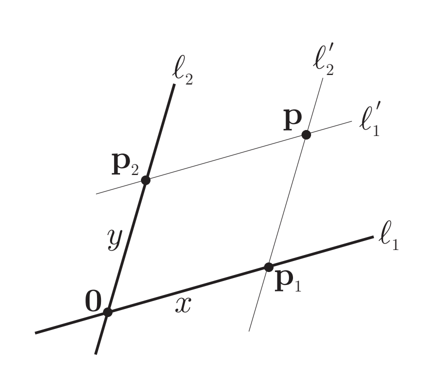
El primer paso es notar que los puntos de una recta $\ell \subset \mathbb{E}^2$ corresponden biunívocamente a los *números reales*, que se denotan por $\mathbb{R}$. Escogemos un punto $\mathbf{o} \in \ell$ que se llamará el *origen* y que corresponderá al número cero. El origen parte a la recta en dos mitades; a una de ellas se le asocian los números positivos de acuerdo con su distancia al origen, y a la otra mitad, los números negativos. Así, a cada número real x se le asocia un punto $\mathbf{p} \in \ell$, y a cada punto en ℓ le corresponde un número real.

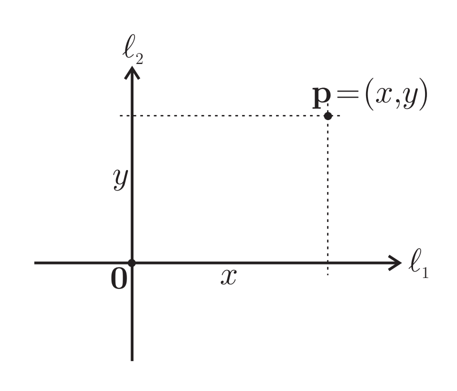


[1]La notación básica de conjuntos se establece e introduce brevemente en el Apéndice.

De esta identificación natural surge el nombre de “recta de los números reales” y todo el desarrollo ulterior de la Geometría Analítica, e inclusive del Cáculo. Los números tienen un significado geométrico (los griegos lo sabían bien al entenderlos como distancias), y entonces problemas geométricos (y físicos) pueden atacarse y entenderse manipulando números.

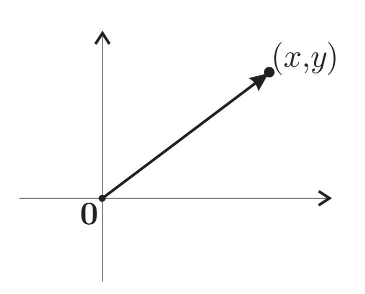
El segundo paso para identificar los puntos del plano euclidiano con parejas de números hace uso esencial del quinto postulado. Tómense dos rectas ℓ_1 y ℓ_2 en el plano $\mathbb{E}^2$ que se intersecten en un punto $\mathbf{o}$ que será el *origen*. Orientamos las rectas para que sus puntos correspondan a números reales, como arriba. Entonces a cualquier punto $\mathbf{p} \in \mathbb{E}^2$ se le puede hacer corresponder una pareja de números de la siguiente forma. Por el Quinto, existe una única recta ℓ_1' que pasa por $\mathbf{p}$ y es paralela a ℓ_1; y análogamente, existe una única recta ℓ_2' que pasa por $\mathbf{p}$ y es paralela a ℓ_2. Las intersecciones $\ell_1 \cap \ell_2'$ y $\ell_2 \cap \ell_1'$ determinan los puntos $\mathbf{p}_1 \in \ell_1$ y $\mathbf{p}_2 \in \ell_2$, respectivamente, y por tanto, determinan dos números reales x y y; es decir, una pareja ordenada (x, y). Y al revés, dada la pareja de números (x, y), tómense $\mathbf{p}_1$ como el punto que está sobre ℓ_1 a distancia x de $\mathbf{o}$, y $\mathbf{p}_2$ como el punto que está sobre ℓ_2 a distancia y de $\mathbf{o}$ (tomando en cuenta signos, por supuesto). Sea ℓ_1' la recta que pasa por $\mathbf{p}_2$ paralela a ℓ_1 y sea ℓ_2' la recta que pasa por $\mathbf{p}_1$ paralela a ℓ_2 (¡acabamos de usar de nuevo al Quinto!). La intersección $\ell_1' \cap \ell_2'$ es el punto $\mathbf{p}$ que corresponde a la pareja (x, y).





La generalidad con que hicimos el razonamiento anterior fue para hacer explícito el uso del quinto postulado en el razonamiento clásico de Descartes, pero no conviene dejarlo tan ambiguo. Es costumbre tomar a las rectas ℓ_1 y ℓ_2 ortogonales: a ℓ_1 horizontal con su dirección positiva a la derecha (como vamos leyendo), conocida tradicionalmente como el *eje de las* x ; y a ℓ_2 vertical con dirección positiva hacia arriba, el famoso *eje de las* y . No sólo es por costumbre que se utiliza esta convención, sino también porque simplifica el álgebra asociada a la geometría clásica. Por ejemplo, permitirá usar el teorema de Pitágoras para calcular facilmente las distancias a partir de las coordenadas. Pero esto se verá más adelante.

En resumen, al fijar los ejes coordenados, a cada pareja de números (x, y) le corresponde un punto $\mathbf{p} \in \mathbb{E}^2$, pero iremos más allá y los identificaremos diciendo $\mathbf{p} = (x, y)$. Además, se le puede hacer corresponder la flecha (segmento de recta dirigido llamado *vector*) que "nace" en el origen y termina en el punto. Así, las siguientes interpretaciones son equivalentes y se usarán de manera indistinta a lo largo de todo el libro.



a. Un punto en el plano.
b. Una pareja ordenada de números reales.
c. Un vector que va del origen al punto.

Nótese que si bien en este texto las palabras *punto* y *vector* son equivalentes, tienen diferentes connotaciones. Mientras que un punto es pensado como un lugar (una posición) en el espacio, un vector es pensado como un segmento de línea dirigido de un punto a otro.

El conjunto de todas las parejas ordenadas de números se denota por $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. En este caso el exponente 2 sí se refiere a exponenciación, pero de conjuntos, lo que se conoce ahora como el *producto cartesiano* en honor de Descartes (véase el Apéndice A); aunque la usanza común es leerlo "$\mathbb{R}$ dos".

1.2.1 Geometría analítica

La idea genialmente simple $\mathbb{R}^2 = \mathbb{E}^2$ (es decir, las parejas ordenadas de números reales se identifican naturalmente con los puntos del plano euclidiano) logra que converjan las aguas del álgebra y la geometría. A partir de Descartes se abre la posibilidad de reconstruir la geometría de los griegos sobre la base de nuestra intuición númerica (básicamente la de sumar y multiplicar). Y a su vez, la luz de la geometría baña con significados y problemas al álgebra. Se hermanan, se apoyan, se entrelazan: nace la geometría analítica.

De aquí en adelante, salvo en contadas ocasiones donde sea inevitable y como comentarios, abandonaremos la línea del desarrollo histórico. De hecho, nuestro tratamiento algebraico está muy lejos del original de Descartes, pues usa ideas desarrolladas en el siglo XIX. Pero antes de entrar de lleno a él, vale la pena enfatizar que no estamos abandonando el método axiomático iniciado por Euclides, simplemente cambiaremos de conjunto de axiomas: ahora nos basaremos en los axiomas de los números reales (ver la sección 1.3.1). De tal manera que los objetos primarios de Euclides (línea, segmento, distancia, ángulo, etc.) serán definiciones (por ejemplo, *punto* ya es "pareja ordenada de números" y *plano* es el conjunto de todos los puntos), y los cinco postulados serán teoremas que hay que demostrar. Nuestros objetos primarios básicos serán ahora los números reales y nuestra intuición geométrica primordial es que forman una recta: La Recta Real, $\mathbb{R}$. Sin embargo, la motivación básica para las definiciones sigue siendo la intuición geométrica desarrollada por los griegos. No queremos hacernos los occisos como si no conocieramos la geometría

griega; simplemente la llevaremos a nuevos horizontes con el apoyo del álgebra —el enfoque analítico— y para ello tendremos que reconstruirla.

EJERCICIO 1.3 Encuentra diferentes puntos en el plano dados por sus coordenadas (por ejemplo, el $(3, 2)$, el $(-1, 5)$, el $(0, -4)$, el $(-1, -2)$, el $(1/2, -3/2)$).

EJERCICIO 1.4 Identifica los *cuadrantes* donde las parejas tienen signos determinados.

EJERCICIO 1.5 ¿Cúales son las coordenadas de los vértices de un cuadrado de lado 2, centrado en el origen y con lados paralelos a los ejes?

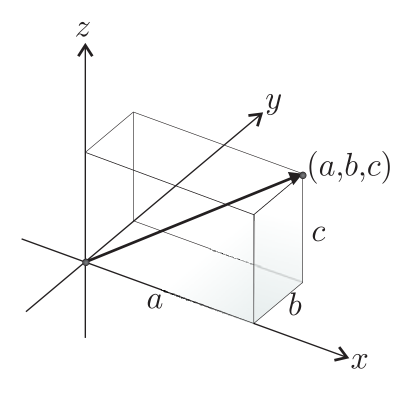
EJERCICIO 1.6 ¿Cúales son las coordenadas de los vértices del octágono regular que incluye como vértices a los del cuadrado anterior? (*Tienes que usar el Teorema de Pitágoras.*)

EJERCICIO 1.7 ¿Puedes dar las coordenadas de los vértices de un triángulo equilátero centrado en el origen? Dibújalo. (*Usa de nuevo el Teorema de Pitágoras, en un triángulo rectángulo con hipotenusa* 2 *y un cateto* 1).

1.2.2 El espacio de dimensión n

La posibilidad de extender con paso firme la geometría euclidiana a más de dos dimensiones es una de las aportaciones de mayor profundidad del método cartesiano.

Aunque nos hayamos puesto formales en la sección anterior para demostrar una correspondencia natural entre $\mathbb{E}^2$ y $\mathbb{R}^2$, intuitivamente es muy simple. Al fijar los dos ejes coordenados —las dos direcciones preferidas— se llega de manera inequívoca a cualquier punto en el plano dando las distancias que hay que recorrer en esas direcciones para llegar a él. En el espacio, habrá que fijar tres ejes coordenados y dar tres números. Los dos primeros dan un punto en el plano (que podemos pensar horizontal, como el piso), y el tercero nos da la altura (que puede ser positiva o negativa). Si denotamos por $\mathbb{R}^3$ al conjunto de todas las ternas (x, y, z) de números reales, como éstas corresponden a puntos en el espacio una vez que se fijan los tres ejes (el eje vertical se conoce como *eje* z), podemos definir el espacio euclidiano de tres dimensiones como $\mathbb{R}^3$, sin preocuparnos de axiomas. ¿Y por qué pararse en 3? ¿o en 4?



Dado un número natural $n \in \{1, 2, 3, \dots\}$, denotamos por $\mathbb{R}^n$ al conjunto de todas las n-adas (léase "eneadas") ordenadas de números reales $(x_1, x_2, \dots, x_n)$. Formalmente,

$$\mathbb{R}^n := \{(x_1, x_2, \dots, x_n) \mid x_i \in \mathbb{R}, i = 1, 2, \dots, n\}.$$

Para valores pequeños de n, se pueden usar letras x, y, z e inclusive w para denotar las coordenadas; pero para n general esto es imposible y no nos queda más que

usar los subíndices. Así, tenemos un conjunto bien definido, $\mathbb{R}^n$, al que podemos llamar "*espacio euclidiano de dimensión* n" y hacer geometría en él. A una n-ada $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ se le llama *vector* o *punto*.

El estudiante que se sienta incómodo con esto de "muchas dimensiones", puede sustituir (pensar en) un 2 o un 3 siempre que se se use n y referirse al plano o al espacio tridimensional que habitamos para su intuición. No pretendemos en este libro estudiar la geometría de estos espacios de muchas dimensiones. Nos concentraremos en dimensiones 2 y 3, así que puede pensarse la n como algo que puede tomar valores 2 o 3 y que resulta muy útil para hablar de los dos al mismo tiempo pero en singular. Sin embargo, es importante que el estudiante tenga una visión amplia y pueda preguntarse en cada momento ¿qué pasaría en 4 o más dimensiones? Como verá, y a veces señalaremos explícitamente, algunas cosas son muy fáciles de generalizar, y otras no tanto. Por el momento, basta con volver a enfatizar el amplísimo mundo que abrió el método de las coordenadas cartesianas para las matemáticas y la ciencia.

Vale la pena en este momento puntualizar las convenciones de notación que utilizaremos en este libro. A los números reales los denotamos por letras simples, por ejemplo x, y, z, o bien a, b y c o bien t, r y s; e inclusive conviene a veces utilizar letras griegas, α (alfa), β (beta) y γ (gamma), o λ (lambda) y μ (mu) (por alguna razón hemos hecho costumbre de utilizarlas siempre en pequeñas familias). Por su parte, a los puntos o vectores los denotamos por letras en negritas, por ejemplo, $\mathbf{p}$ y $\mathbf{q}$, para señalar que estamos pensando en su caracter de puntos, o bien, usamos $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ o $\mathbf{d}$ (acrónimo elegante que usaremos para "dirección") y $\mathbf{n}$ (acrónimo para "normal") para subrayar su papel vectorial. Por supuesto, $\mathbf{x}$, $\mathbf{y}$ y $\mathbf{z}$ siempre son buenos comodines para vectores (puntos o eneadas) variables o incógnitos, que no deben que confundirse con x, y, z, que son variables reales o numéricas.

1.3 El espacio vectorial $\mathbb{R}^2$

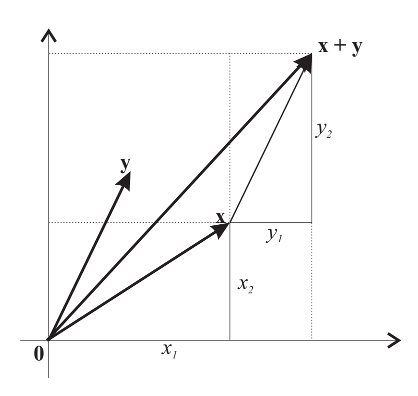
En esta sección se introduce la herramienta algebraica básica para hacer geometría con parejas, ternas o n-adas de números. Y de nuevo la idea central es muy simple: así como los números se pueden sumar y multiplicar, también los vectores tienen sus operacioncitas. Lo único que suponemos de los números reales es que sabemos o, mejor aún, que "saben ellos" sumarse y multiplicarse; a partir de ello extenderemos estas nociones a los vectores.

Definición 1.3.1 Dados dos vectores $\mathbf{u} = (x, y)$ y $\mathbf{v} = (x', y')$ en $\mathbb{R}^2$, definimos su *suma vectorial*, o simplemente su *suma*, como el vector $\mathbf{u} + \mathbf{v}$ que resulta de sumar coordenada a coordenada:

$$\mathbf{u} + \mathbf{v} := (x + x', y + y'),$$

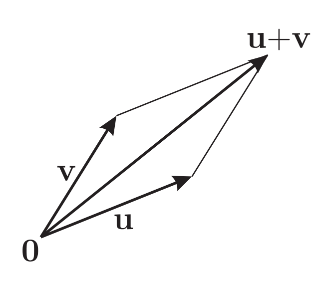
es decir,

$$(x, y) + (x', y') := (x + x', y + y').$$

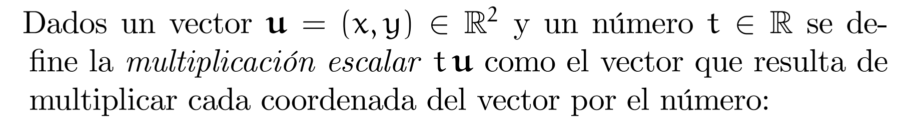


Nótese que en cada coordenada, la suma que se usa es la suma usual de números reales. Así que al signo "+" de la suma le hemos ampliado el espectro de "saber sumar números" a "saber sumar vectores"; pero con una receta muy simple: "coordenada a coordenada". Por ejemplo,

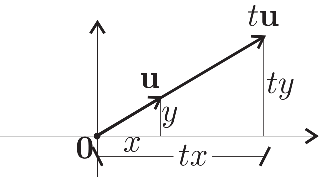
$$(3,2) + (-1,1) = (2,3).$$



La suma vectorial corresponde geométricamente a la regla del paralelogramo usada para encontrar la resultante de dos vectores. Esto es, se consideran los vectores como segmentos dirigidos que salen del origen, generan entonces un paralelogramo y el vector que va del origen a la otra esquina es la suma. También se puede pensar como la acción de dibujar un vector tras otro, pensando que los vectores son segmentos dirigidos que pueden moverse paralelos a sí mismos.

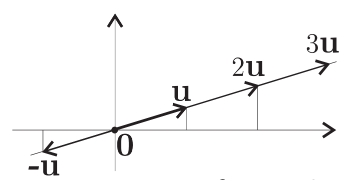


Definición 1.3.2 Dados un vector $\mathbf{u} = (x, y) \in \mathbb{R}^2$ y un número $t \in \mathbb{R}$ se define la *multiplicación escalar* $t\,\mathbf{u}$ como el vector que resulta de multiplicar cada coordenada del vector por el número:



$$t\,\mathbf{u} := (t\,x, t\,y).$$

Nótese que en cada coordenada, la multiplicación que se usa es la de los números reales.



La multiplicación escalar corresponde a la dilatación o contracción, y posiblemente al cambio de dirección, de un vector. Veamos. Está claro que

$$2\mathbf{u} = (2x, 2y) = (x + x, y + y) = \mathbf{u} + \mathbf{u},$$

así que $2\mathbf{u}$ es el vector "$\mathbf{u}$ seguido de $\mathbf{u}$" o bien, "$\mathbf{u}$ dilatado a su doble". De la misma manera que $3\mathbf{u}$ es un vector que apunta en la misma dirección pero de tres veces su tamaño. O bien, es fácil deducir que $(\frac{1}{2})\mathbf{u}$, como punto, está justo a la mitad del camino del origen $0 = (0,0)$ a $\mathbf{u}$. Así que $t\,\mathbf{u}$ para $t > 1$ es, estrictamente hablando, una *dilatación* de $\mathbf{u}$, y para $0 < t < 1$, una *contracción* del mismo. Por último, para $t < 0$, $t\,\mathbf{u}$ apunta en la dirección contraria ya que, en particular, $(-1)\mathbf{u} =: -\mathbf{u}$ es el vector que, como segmento dirigido, va del punto $\mathbf{u}$ al 0 (puesto que $\mathbf{u} + (-\mathbf{u}) = 0$) y el resto se obtiene como dilataciones o contracciones de $-\mathbf{u}$.

No está de más insistir en una diferencia esencial entre las dos operaciones que hemos definido. Si bien la suma vectorial es una operación que de dos ejemplares de la misma especie (vectores) nos da otro de ellos, la multiplicación escalar involucra a dos objetos de diferente índole, por un lado el "escalar", un número real, y por el otro un vector, lo que da como resultado un nuevo vector. Los vectores no se multiplican (por el momento), sólo los escalares (los números reales) saben multiplicarlos (pegarles, podría decirse) y les cambian con ello su "escala".

EJERCICIO 1.8 Sean $v_1 = (2,3)$, $v_2 = (-1,2)$, $v_3 = (3,-1)$ y $v_4 = (1,-4)$.
i) Calcula y dibuja: $2v_1 - 3v_2$; $2(v_3 - v_4) - v_3 + 2v_4$; $2v_1 - 3v_3 + 2v_4$.
ii) ¿Qué vector $x \in \mathbb{R}^2$ cumple que $2v_1 + x = 3v_2; 3v_3 - 2x = v_4 + x$?
iii) ¿Puedes encontrar $r, s \in \mathbb{R}$ tales que $r v_2 + s v_3 = v_4$?

EJERCICIO 1.9 Dibuja el origen y tres vectores cualesquiera u, v y w en un papel. Con un par de escuadras encuentra los vectores $u + v$, $v + w$ y $w + u$.

EJERCICIO 1.10 Dibuja el origen y dos vectores cualesquiera u, v en un papel. Con regla (escuadras) y compás, encuentra los puntos $(1/2)\,u + (1/2)\,v$, $2u - v$, $3u - 2v$, $2v - u$. ¿Resultan colineales?

EJERCICIO 1.11 Describe el subconjunto de $\mathbb{R}^2$, o *lugar geométrico*, definido por $\mathcal{L}_v := \{t\,v \mid t \in \mathbb{R}\}$.

Estas definiciones se extienden a $\mathbb{R}^n$ de manera natural. Dados dos vectores $x = (x_1, \ldots, x_n)$ y $y = (y_1, \ldots, y_n)$ en $\mathbb{R}^n$ y un número real $t \in \mathbb{R}$, la *suma vectorial* (o simplemente la *suma*) y el *producto (o multiplicación) escalar* se definen como sigue:

$$x + y := (x_1 + y_1, \ldots, x_n + y_n),$$
$$t\,x := (t\,x_1, \ldots, t\,x_n).$$

Es decir, la suma de dos vectores (con el mismo número de coordenadas) se obtiene sumando coordenada a coordenada y el producto por un escalar (un número) se obtiene multiplicando cada coordenada por ese número.

Las propiedades básicas de la suma vectorial y la multiplicación escalar se reúnen en el siguiente teorema, donde el vector $0 = (0, \ldots, 0)$ es llamado *vector cero* que corresponde al *origen*, y, para cada $x \in \mathbb{R}^n$, el vector $-x := (-1)x$ se llama *inverso aditivo* de x.

Teorema 1.2 *Para todos los vectores* x, y, $z \in \mathbb{R}^n$ *y para todos los números* s, $t \in \mathbb{R}$ *se cumple que:*

i) $(x + y) + z = x + (y + z)$
ii) $x + y = y + x$
iii) $x + 0 = x$
iv) $x + (-x) = 0$
v) $s(t\,x) = (st)x$
vi) $1x = x$
vii) $t\,(x + y) = t\,x + t\,y$
viii) $(s + t)\,x = s\,x + t\,x$

1.3.1 ¿Teorema o axiomas?

Antes de pensar en demostrar el teorema anterior, vale la pena reflexionar un poco sobre su carácter pues está muy cerca de ser un conjunto de axiomas y es sutil qué

quiere decir eso de demostrarlo.

Primero, debemos observar que $\mathbb{R}^n$ tiene sentido para $n = 1$, y es simplemente una manera rimbombante de referirse a los números reales. Así que del teorema, siendo $n = 1$ (es decir, si x, y, z están en $\mathbb{R}$), se obtiene parte de los *axiomas de los números reales.* Esto es, cada enunciado es una de las reglas elementales de la suma y la multiplicación que conocemos desde pequeños. Las propiedades **(i)** a **(iv)** son los axiomas que hacen a $\mathbb{R}$, con la operación suma, lo que se conoce como un "*grupo conmutativo*": **(i)** dice que la suma es "*asociativa*", **(ii)** que es "*conmutativa*", **(iii)** que el 0 —o bien, el número 0— es su "*neutro*" (aditivo) y **(iv)** que todo elemento tiene "*inverso*" (aditivo). Por su parte, **(v)** y **(vi)** nos dicen que la multiplicación es asociativa y que tiene un neutro (multiplicativo), el 1; pero nos faltaría que es conmutativa y que tiene inversos (multiplicativos). Entonces tendríamos que añadir:

ix) $t\,s = s\,t$
x) Si $t \neq 0$, existe t^{-1} tal que $t\,t^{-1} = 1$

para obtener que $\mathbb{R} - \{0\}$ (los reales sin el 0) son un grupo conmutativo con la multiplicación. Finalmente, **(vii)** y **(viii)** ya dicen lo mismo (en virtud de **(ix)**), que las dos operaciones se *distribuyen.* Estos axiomas definen lo que se llama un *campo.*

Obsérvese que en el caso general ($n > 1$) **(ix)** y **(x)** ni siquiera tienen sentido; pues sólo cuando $n = 1$ la multiplicación escalar involucra a seres de la misma especie. En resumen, para el caso $n = 1$, el teorema es un subconjuto de los axiomas que definen las operaciones de los números reales. No hay nada que demostrar, pues son parte de los axiomas básicos que vamos a usar. El resto de los axiomas que definen los números reales se refieren a su orden y, para que el libro quede autocontenido, de una vez los enunciamos. Los números reales tienen una *relación de orden*, denotada por $\leq$ y que se lee "*menor o igual que*" que cumple, para $a, b, c, d \in \mathbb{R}$, con:

Oi) $a \leq b$ o $b \leq a$ (es un orden total)
Oii) $a \leq b$ y $b \leq a$ $\Leftrightarrow$ $a = b$
Oiii) $a \leq b$ y $b \leq c$ $\Rightarrow$ $a \leq c$ (es transitivo)
Oiv) $a \leq b$ y $c \leq d$ $\Rightarrow$ $a + c \leq b + d$
Ov) $a \leq b$ y $0 \leq c$ $\Rightarrow$ $ac \leq bc$

Además, los números reales cumplen con el *axioma del supremo* que, intuitivamente, dice que la recta numérica no tiene "hoyos", que los números reales forman un continuo. Pero este axioma, fundamental para el cálculo pues hace posible formalizar lo "infinitesimal", no se usa en este libro, así que lo dejamos de lado.

Regresando al Teorema 1.2, para $n > 1$ la cosa es sutilmente diferente. Nosotros definimos la suma vectorial, y al ser algo nuevo sí tenemos que verificar que cumple las propiedades requeridas. Vale la pena introducir a Dios, el de las matemáticas no el de las religiones, para que quede claro. Dios nos da los números reales con la suma y la multiplicación, de alguna manera nos "ilumina" y de golpe: ahí están, con todo y sus axiomas. Ahora, nosotros —simples mortales— osamos definir la suma vectorial

y la multiplicación de escalares a vectores, pero Dios ya no nos asegura nada, él ya hizo lo suyo: nos toca demostrarlo a nosotros.

Demostración. (del Teorema 1.2) Formalmente sólo nos interesa demostrar el teorema para $n = 2, 3$, pero esencialmente es lo mismo para cualquier n. La demostración de cada inciso es muy simple, tanto así que hasta confunde, y consiste en aplicar el axioma correspondiente que cumplen los números reales coordenada a coordenada y la definición de las operaciones involucradas. Sería muy tedioso, y aportaría muy poco al lector, demostrar los ocho incisos, así que sólo demostraremos con detalle el primero para $\mathbb{R}^2$, dejando los demás y el caso general como ejercicios.

i) Sean $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in \mathbb{R}^2$, entonces $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2)$ y $\mathbf{z} = (z_1, z_2)$, donde cada x_i, y_i y z_i $(i = 1, 2)$ son números reales (nótese que conviene usar subíndices con la misma letra que en negritas denota al vector). Tenemos entonces, a partir de la definición de suma vectorial, que

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = (x_1 + y_1, x_2 + y_2) + (z_1, z_2)$$

y usando nuevamente la definición se obtiene que esta última expresión es

$$= ((x_1 + y_1) + z_1, (x_2 + y_2) + z_2).$$

Luego, como *la suma de números reales es asociativa* (el axioma correspondiente de los números reales usado coordenada a coordenada) se sigue

$$= (x_1 + (y_1 + z_1), x_2 + (y_2 + z_2)).$$

Y finalmente, usando dos veces la definición de suma vectorial, se obtiene

$$\begin{aligned} &= (x_1, x_2) + (y_1 + z_1, y_2 + z_2) \\ &= (x_1, x_2) + ((y_1, y_2) + (z_1, z_2)) \\ &= \mathbf{x} + (\mathbf{y} + \mathbf{z}), \end{aligned}$$

lo que demuestra que $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$; y entonces tiene sentido escribir $\mathbf{x} + \mathbf{y} + \mathbf{z}$. □

Se conoce como *espacio vectorial* a un conjunto en el que están definidas dos operaciones (suma vectorial y multiplicación escalar) que cumplen con las ocho propiedades del Teorema 1.2. De tal manera que este teorema puede parafrasearse "$\mathbb{R}^n$ es un espacio vectorial". ¿Podría el lector mencionar un espacio vectorial distinto de $\mathbb{R}^n$?

EJERCICIO 1.12 Usando los axiomas de los números reales, así como las definiciones de suma vectorial y producto escalar, demuestra algunos incisos del Teorema 1.2 para el caso $n = 2$ y $n = 3$. ¿Verdad que lo único que cambia es la longitud de los vectores, pero los argumentos son exactamente los mismos? Demuestra (aunque sea mentalmente) el caso general.

EJERCICIO 1.13 Demuestra que si $a \leq 0$ entonces $0 \leq -a$. (Usa el axioma (**Oiv**).)

EJERCICIO 1.14 Observa que los axiomas de orden no dicen nada sobre cómo están relacionados el 0 y el 1, pero se pueden deducir de ellos. Supón que $1 \leq 0$, usando el ejercicio anterior y los axiomas (**Ov**) y (**Oii**), demuestra que entonces $-1 = 0$; como esto no es cierto se debe cumplir la otra posibilidad por (**Oi**), es decir, que $0 \leq 1$.

EJERCICIO 1.15 Demuestra que si $a \leq b$ y $c \leq 0$ entonces $ac \geq bc$ (donde $\geq$ significa "es *mayor o igual*").

Ejemplo. Como corolario de este teorema se tiene que el "algebra" simple a la que estamos acostumbrados con números también vale con los vectores. Por ejemplo, en el ejercicio (1.8.ii.b) se pedía encontrar el vector $\mathbf{x}$ tal que

$$3\mathbf{v}_3 - 2\mathbf{x} = \mathbf{v}_4 + \mathbf{x}.$$

Hagámoslo. Se vale pasar con signo menos del otro lado de la ecuación (sumando el inverso aditivo a ambos lados), y por tanto esta ecuación equivale a

$$-3\mathbf{x} = \mathbf{v}_4 - 3\mathbf{v}_3$$

y multiplicando por $-1/3$ se obtiene que

$$\mathbf{x} = \mathbf{v}_3 - (1/3)\mathbf{v}_4.$$

Ya nada más falta sustituir por los valores dados. Es probable que el estudiante haya hecho algo similar y qué bueno: tenía la intuición correcta. Pero hay que tener cuidado, así como con los números no se puede dividir entre cero, no se le vaya a ocurrir tratar de ¡dividir entre un vector! Aunque a veces se pueden "cancelar" (ver Ej. 1.17). Para tal efecto, el siguiente lema será muy usado y vale la pena verlo en detalle.

Lema 1.3 *Si* $\mathbf{x} \in \mathbb{R}^2$ *y* $t \in \mathbb{R}$ *son tales que* $t\mathbf{x} = \mathbf{0}$ *entonces* $t = 0$ *o* $\mathbf{x} = \mathbf{0}$.

Demostración. Suponiendo que $t \neq 0$, hay que demostrar que $\mathbf{x} = \mathbf{0}$ para completar el lema. Pero entonces t tiene inverso multiplicativo y podemos multiplicar por t^{-1} ambos lados de la ecuación $t\mathbf{x} = \mathbf{0}$ para obtener (usando (**v**) y (**vi**)) que $\mathbf{x} = \left(t^{-1}\right)\mathbf{0} = \mathbf{0}$ por la definición del vector nulo $\mathbf{0} = (0, 0)$. □

EJERCICIO 1.16 Demuestra que si $\mathbf{x} \in \mathbb{R}^3$ y $t \in \mathbb{R}$ son tales que $t\mathbf{x} = \mathbf{0}$ entonces $t = 0$ o $\mathbf{x} = \mathbf{0}$. ¿Y para $\mathbb{R}^n$?

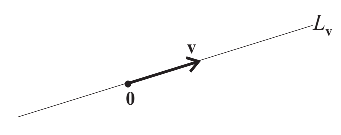
EJERCICIO 1.17 Demuestra que si $\mathbf{x} \in \mathbb{R}^n$ es distinto de $\mathbf{0}$, y $t, s \in \mathbb{R}$ son tales que $t\mathbf{x} = s\mathbf{x}$, entonces $t = s$. (Es decir, si $\mathbf{x} \neq \mathbf{0}$ está permitido "cancelarlo" aunque sea vector.)

1.4 Líneas

En el estudio de la geometría clásica, las líneas o rectas son subconjuntos básicos descritos por los axiomas; se les reconoce intuitivamente y se parte de ellas para construir lo demás. Con el método cartesiano, esto no es necesario. Las líneas se pueden definir o construir, correspondiendo a nuestra noción intuitiva de ellas —que no varía en nada de la de los griegos—, para después ver que efectivamente cumplen con los axiomas que antes se les asociaban. En esta sección, definiremos las rectas y veremos que cumplen los axiomas primero y tercero de Euclides.

Nuestra definición de recta estará basada en la intuición física de una partícula que viaja en movimiento rectilíneo uniforme; es decir, con velocidad fija y en la misma dirección. Estas dos nociones (velocidad como magnitud y dirección) se han amalgamado en el concepto de vector.

Recordemos que en el Ejercicio 1.11, se pide describir el conjunto $\mathcal{L}_{\mathbf{v}} := \{t\mathbf{v} \in \mathbb{R}^2 \mid t \in \mathbb{R}\}$ donde $\mathbf{v} \in \mathbb{R}^2$. Debía ser claro que si $\mathbf{v} \neq \mathbf{0}$, entonces $\mathcal{L}_{\mathbf{v}}$, al constar de todos los múltiplos escalares del vector $\mathbf{v}$, se dibuja como una recta que pasa por el origen (pues $\mathbf{0} = 0\mathbf{v}$) con la dirección de $\mathbf{v}$. Si consideramos la variable t como el tiempo, la función $\varphi(t) = t\mathbf{v}$ describe el movimiento rectilíneo uniforme de una partícula que viene desde tiempo infinito negativo, pasa por el origen $\mathbf{0}$ en el tiempo $t = 0$ (llamada la posición inicial), y seguirá por siempre con velocidad constante $\mathbf{v}$.

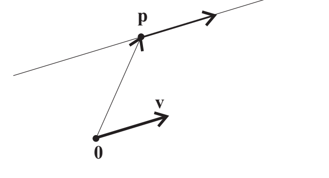


Pero por el momento queremos pensar en las rectas como conjuntos (en el capítulo siguiente estudiaremos más a profundidad las funciones). Los conjuntos $\mathcal{L}_{\mathbf{v}}$, al rotar $\mathbf{v}$, nos dan las rectas por el origen, y para obtener otras, bastará "empujarlas" fuera del origen (o bien, arrancar el movimiento con otra posición inicial).

Definición 1.4.1 Dados un punto $\mathbf{p}$ y un vector $\mathbf{v} \neq \mathbf{0}$, la *recta que pasa por* $\mathbf{p}$ con dirección $\mathbf{v}$ es el conjunto:

$$\ell := \{\mathbf{p} + t\mathbf{v} \mid t \in \mathbb{R}\}. \tag{1.1}$$

Una *recta* o *línea* en $\mathbb{R}^2$ es un subconjunto que tiene, para algún $\mathbf{p}$ y $\mathbf{v} \neq \mathbf{0}$, la descripción anterior.



A esta forma de definir una recta se le conoce como *representación o expresión paramétrica*. Esta representación trae consigo una función entre los números reales y la recta:

$$\begin{aligned} \varphi &: \mathbb{R} \to \mathbb{R}^2 \\ \varphi(t) &= \mathbf{p} + t\mathbf{v}. \end{aligned} \tag{1.2}$$

De hecho, esta función define una biyección entre $\mathbb{R}$ y ℓ. Como ℓ se definió por medio del *parámetro* $t \in \mathbb{R}$, es claro que *es* la imagen de la función φ; es decir, φ es sobre. Demostremos (aunque de la intuición física parezca obvio) que es uno-a-uno. Supongamos para esto que $\varphi(t) = \varphi(s)$ para algunos $t, s \in \mathbb{R}$ y habrá que concluir

que $t = s$. Por la definición de φ se tiene

$$\mathbf{p} + t\boldsymbol{v} = \mathbf{p} + s\boldsymbol{v}.$$

De aquí, al sumar el inverso del lado izquierdo a ambos (que equivale a pasarlo de un lado al otro con signo contrario) se tiene

$$\begin{aligned} 0 &= \mathbf{p} + s\boldsymbol{v} - (\mathbf{p} + t\boldsymbol{v}) \\ &= \mathbf{p} + s\boldsymbol{v} - \mathbf{p} - t\boldsymbol{v}) \\ &= (\mathbf{p} - \mathbf{p}) + (s\boldsymbol{v} - t\boldsymbol{v}) \\ &= 0 + (s - t)\boldsymbol{v} \\ &= (s - t)\,\boldsymbol{v}, \end{aligned}$$

donde hemos usado las propiedades (**i**), (**ii**), (**iii**), (**iv**) y (**viii**) del Teorema 1.2.[2] Del Lema 1.3, se concluye que $s - t = 0$ o bien que $\boldsymbol{v} = 0$. Pero por hipótesis $\boldsymbol{v} \neq 0$ (esto es la esencia), así que no queda otra que $s = t$. Esto demuestra que φ es inyectiva, y nuestra intuición se corrobora.

Hemos demostrado que cualquier recta está en biyección natural con $\mathbb{R}$, la recta modelo, así que todas las líneas son una “copia” de la recta real abstracta $\mathbb{R}$.

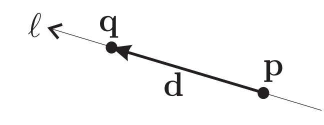
Observación 1.4.1 En la definición 1.4, así como en la demostración, no se usa de manera esencial que $\mathbf{p}$ y $\boldsymbol{v}$ estén en $\mathbb{R}^2$. Sólo se usa que hay una suma vectorial y un producto escalar bien definidos. Así que el punto y el vector podrían estar en $\mathbb{R}^3$ o en $\mathbb{R}^n$. Podemos entonces dar por establecida la noción de recta en cualquier dimensión al cambiar 2 por n en la Definición 1.4.

Observación 1.4.2 A la función 1.2 se le conoce como “*movimiento rectilíneo uniforme*” o “*movimiento inercial*”, pues, como decíamos al principio de la sección, describe la manera en que se mueve una partícula, o un cuerpo, que no está afectada por ninguna fuerza y tiene posición $\mathbf{p}$ y vector velocidad $\boldsymbol{v}$ en el tiempo $t = 0$.

Ya podemos demostrar un resultado clásico, e intuitivamente claro.

Lema 1.4 *Dados dos puntos* $\mathbf{p}$ *y* $\mathbf{q}$ *en* $\mathbb{R}^n$*, existe una recta que pasa por ellos.*

Demostración. Si tuvieramos que $\mathbf{p} = \mathbf{q}$, como hay muchas rectas que pasan por $\mathbf{p}$, ya acabamos. Supongamos entonces que $\mathbf{p} \neq \mathbf{q}$, que es el caso interesante.



Si tomamos a $\mathbf{p}$ como punto base para la recta que buscamos, bastará encontrar un vector que nos lleve de $\mathbf{p}$ a $\mathbf{q}$, para tomarlo como dirección. Éste es la diferencia $\mathbf{q} - \mathbf{p}$, pues claramente

$$\mathbf{p} + (\mathbf{q} - \mathbf{p}) = \mathbf{q}\,,$$

[2] Ésta es la última vez que haremos mención explícita del uso del Teorema 1.2; de aquí en adelante se aplicarán sus propiedades como si fueran lo más natural del mundo. Pero vale la pena que el estudiante se cuestione por algún tiempo qué es lo que se está usando en las manipulaciones algebraicas.

de tal manera que si definimos $\mathbf{d} := \mathbf{q} - \mathbf{p}$, como dirección, la recta

$$\ell = \{\mathbf{p} + t\,\mathbf{d} \mid t \in \mathbb{R}\},$$

(que sí es una recta pues $\mathbf{d} = \mathbf{q} - \mathbf{p} \neq \mathbf{0}$), es la que funciona. Con $t = 0$ obtenemos que $\mathbf{p} \in \ell$, y con $t = 1$ que $\mathbf{q} \in \ell$. □

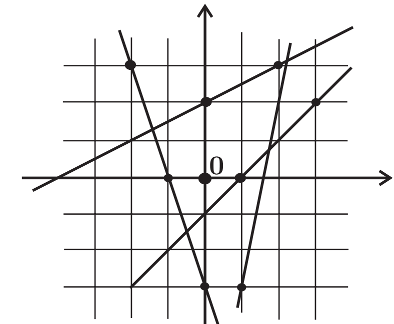
De la demostración del lema, se siguen los postulados I y III de Euclides. Obsérvese que cuando t toma valores entre 0 y 1, se obtienen puntos entre $\mathbf{p}$ y $\mathbf{q}$ (pues el vector direccional $(\mathbf{q} - \mathbf{p})$ se encoge al multiplicarlo por t), así que el *segmento* de $\mathbf{p}$ a $\mathbf{q}$, que denotaremos por $\overline{\mathbf{pq}}$, se debe definir como

$$\overline{\mathbf{pq}} := \{\mathbf{p} + t(\mathbf{q} - \mathbf{p}) \mid 0 \leq t \leq 1\}.$$

Y la recta ℓ que pasa por $\mathbf{p}$ y $\mathbf{q}$ se extiende "indefinidamente" a ambos lados del segmento $\overline{\mathbf{pq}}$; para $t > 1$ del lado de $\mathbf{q}$ y para $t < 0$ del lado de $\mathbf{p}$.

***EJERCICIO 1.18** Aunque lo demostraremos más adelante, es un buen ejercicio demostrar formalmente en este momento que la recta ℓ, cuando $\mathbf{p} \neq \mathbf{q}$, es única.

EJERCICIO 1.19 Encuentra representaciones paramétricas de las rectas en la figura. Observa que su representación paramétrica no es única.



EJERCICIO 1.20 Dibuja las rectas

$$\begin{aligned}
\{(2,3) + t(1,1) &\mid t \in \mathbb{R}\};\\
\{(-1,0) + s(2,1) &\mid s \in \mathbb{R}\};\\
\{(0,-2) + (-r,2r) &\mid r \in \mathbb{R}\};\\
\{(t-1,-2t) &\mid t \in \mathbb{R}\}.
\end{aligned}$$

EJERCICIO 1.21 Exhibe representaciones paramétricas para las seis rectas que pasan por dos de los siguientes puntos en $\mathbb{R}^2$: $\mathbf{p}_1 = (2,4)$, $\mathbf{p}_2 = (-1,2)$, $\mathbf{p}_3 = (-1,-1)$ y $\mathbf{p}_4 = (3,-1)$.

EJERCICIO 1.22 Muestra representaciones paramétricas para las tres rectas que genera el triángulo en $\mathbb{R}^3$: $\mathbf{q}_1 = (2,1,2)$, $\mathbf{q}_2 = (-1,1,-1)$, $\mathbf{q}_3 = (-1,-2,-1)$.

EJERCICIO 1.23 Si una partícula viaja de $2\mathbf{q}_1$ a $3\mathbf{q}_3$ en movimiento inercial y tarda cuatro unidades de tiempo en llegar. ¿Cuál es su vector velocidad?

EJERCICIO 1.24 Dados dos vectores $\mathbf{u}$ y $\mathbf{v}$ en $\mathbb{R}^n$, el *paralelogramo* que definen tiene como vértices los puntos $\mathbf{0}, \mathbf{u}, \mathbf{v}$ y $\mathbf{u} + \mathbf{v}$ (como en la figura). Demuestra que sus *diagonales*, es decir, los segmentos de $\mathbf{0}$ a $\mathbf{u} + \mathbf{v}$ y de $\mathbf{u}$ a $\mathbf{v}$ se intersectan en su punto medio.

1.4.1 Coordenadas baricéntricas

Todavía podemos exprimirle más información a la demostración del Lema 1.4. Veremos cómo se escriben, en términos de $\mathbf{p}$ y $\mathbf{q}$, los puntos de su recta, y que esto tiene que ver con la clásica ley de las palancas. Además, de aquí surgirá una demostración muy simple del teorema clásico de concurrencia de medianas en un triángulo.

Supongamos que $\mathbf{p}$ y $\mathbf{q}$ son dos puntos distintos del plano. Para cualquier $t \in \mathbb{R}$, se tiene que

$$\begin{aligned} \mathbf{p} + t(\mathbf{q} - \mathbf{p}) &= \mathbf{p} + t\,\mathbf{q} - t\,\mathbf{p} \\ &= (1 - t)\mathbf{p} + t\,\mathbf{q} \end{aligned}$$

al reagrupar los términos. Y esta última expresión, a su vez, se puede reescribir como

$$s\,\mathbf{p} + t\,\mathbf{q} \quad \text{con} \quad s + t = 1,$$

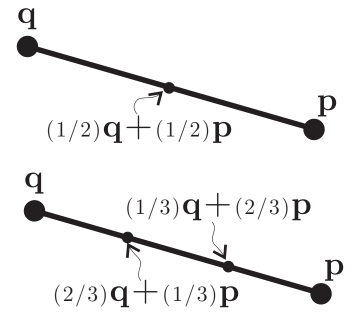
donde hemos introducido la nueva variable $s = 1 - t$. De lo anterior se deduce que la recta ℓ que pasa por $\mathbf{p}$ y $\mathbf{q}$ puede también describirse como

$$\ell = \{s\,\mathbf{p} + t\,\mathbf{q} \mid s + t = 1\},$$

que es una *expresión baricéntrica* de ℓ. A los números s, t (con $s + t = 1$, insistimos) se les conoce como *coordenadas baricéntricas* del punto $\mathbf{x} = s\,\mathbf{p} + t\,\mathbf{q}$ con respecto a $\mathbf{p}$ y $\mathbf{q}$.

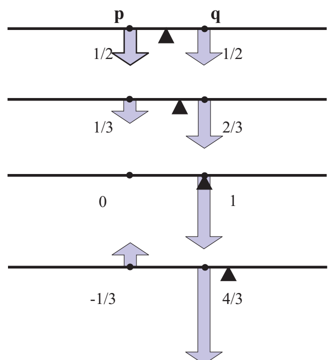
Las coordenadas baricéntricas tienen la ventaja de que ya no distinguen entre los dos puntos. Como lo escribíamos antes —sólo con t—, $\mathbf{p}$ era el punto base ("el que la hace de 0") y luego $\mathbf{q}$ era hacia donde íbamos ("el que la hace de 1"), y tenían un papel distinto. Ahora no hay preferencia por ninguno de los dos; las coordenadas baricéntricas son "democráticas". Más aún, al expresar una recta con coordenadas baricéntricas no le damos una dirección preferida. Se usan simultáneamente los parámetros naturales para las dos direcciones (de $\mathbf{p}$ a $\mathbf{q}$ y de $\mathbf{q}$ a $\mathbf{p}$); pues si $s + t = 1$ entonces

$$\begin{aligned} s\,\mathbf{p} + t\,\mathbf{q} &= \mathbf{p} + t\,(\mathbf{q} - \mathbf{p}) \\ &= \mathbf{q} + s\,(\mathbf{p} - \mathbf{q}). \end{aligned}$$



Nótese que $\mathbf{x} = s\,\mathbf{p} + t\,\mathbf{q}$ está en el segmento entre $\mathbf{p}$ y $\mathbf{q}$ si y sólo si sus dos coordenadas baricéntricas son no negativas, es decir, si y sólo si $t \geq 0$ y $s \geq 0$. Por ejemplo, el punto medio tiene coordenadas baricéntricas $1/2, 1/2$, es $(1/2)\mathbf{p} + (1/2)\mathbf{q}$; y el punto que divide al segmento de $\mathbf{p}$ a $\mathbf{q}$ en la proporción de $2/3$ a $1/3$ se escribe $(1/3)\mathbf{p} + (2/3)\mathbf{q}$, pues se acerca más a $\mathbf{q}$ que a $\mathbf{p}$. La extensión de la recta más allá de $\mathbf{q}$ tiene coordenadas baricéntricas s, t con $s < 0$ (y por lo tanto $t > 1$); así que los puntos de ℓ fuera del segmento de $\mathbf{p}$ a $\mathbf{q}$ tienen alguna coordenada baricéntrica negativa (la correspondiente al punto más lejano).

En términos físicos, podemos pensar la recta por $\mathbf{p}$ y $\mathbf{q}$ como una barra rígida. Si distribuimos una masa (que podemos pensar unitaria, es decir que vale 1) entre estos dos puntos, el punto de equilibrio tiene las coordenadas baricéntricas correspondientes a las masas. Si, por ejemplo, tienen el mismo peso, su punto de equilibrio está en el punto medio. Las masas negativas pueden pensarse como una fuerza hacia arriba y las correspondientes coordenadas baricéntricas nos dan entonces el punto de equilibrio para las palancas.



Veamos ahora un teorema clásico cuya demostración se simplifica enormemente usando coordenadas baricéntricas. Dado un triángulo con vértices $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, se definen las *medianas* como los segmentos que van de un vértice al punto medio del lado opuesto a éste.

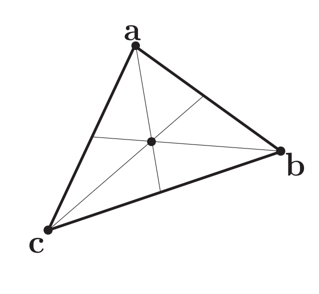
Teorema 1.5 *Dado un triángulo $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, sus tres medianas "concurren" en un punto que las parte en la proporción de 2/3 (del vértice) a 1/3 (del lado opuesto).*

Demostración. Como el punto medio del segmento $\mathbf{b}$, $\mathbf{c}$ es $(\frac{1}{2}\,\mathbf{b} + \frac{1}{2}\,\mathbf{c})$, entonces la mediana por $\mathbf{a}$ es el segmento ...Dibujo

$$\{s\,\mathbf{a} + t\ (\frac{1}{2}\ \mathbf{b} + \frac{1}{2}\ \mathbf{c}) \mid s + t = 1\,,\ s \geq 0\,,\ t \geq 0\},$$

y análogamente se describen las otras dos medianas. Por suerte, el enunciado del teorema nos dice dónde buscar la intersección: el punto que describe en la mediana de $\mathbf{a}$ es precisamente

$$\frac{1}{3}\,\mathbf{a} + \frac{2}{3}\,(\frac{1}{2}\,\mathbf{b} + \frac{1}{2}\,\mathbf{c}) = \frac{1}{3}\,\mathbf{a} + \frac{1}{3}\,\mathbf{b} + \frac{1}{3}\,\mathbf{c}.$$



Entonces, de las igualdades

$$\begin{aligned}\frac{1}{3}\,\mathbf{a} + \frac{1}{3}\,\mathbf{b} + \frac{1}{3}\,\mathbf{c} &= \frac{1}{3}\,\mathbf{a} + \frac{2}{3}\,(\frac{1}{2}\,\mathbf{b} + \frac{1}{2}\,\mathbf{c}) \\ &= \frac{1}{3}\,\mathbf{b} + \frac{2}{3}\,(\frac{1}{2}\,\mathbf{c} + \frac{1}{2}\,\mathbf{a}) \\ &= \frac{1}{3}\,\mathbf{c} + \frac{2}{3}\,(\frac{1}{2}\,\mathbf{a} + \frac{1}{2}\,\mathbf{b})\end{aligned}$$

se deduce que las tres medianas *concurren* en el punto $\frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c})$, es decir, pasan por él. Este punto se llama el *baricentro* del triángulo, y claramente parte a las medianas en la proporción deseada. De nuevo, el baricentro corresponde al "centro de masa" o "punto de equilibrio" del triángulo.

□

EJERCICIO 1.25 Si tienes una barra rígida de un metro y con una fuerza de 10 kg quieres levantar una masa de 40 kg, ¿de dónde debes colgar la masa, estando el punto de apoyo al extremo de tu barra? Haz un dibujo.

EJERCICIO 1.26 Sean $\mathbf{a}$, $\mathbf{b}$ y $\mathbf{c}$ tres puntos distintos entre sí. Demuestra que si $\mathbf{a}$ está en la recta que pasa por $\mathbf{b}$ y $\mathbf{c}$, entonces $\mathbf{b}$ está en la recta por $\mathbf{a}$ y $\mathbf{c}$.

EJERCICIO 1.27 Encuentra el baricentro del triángulo $\mathbf{a} = (2,4)$, $\mathbf{b} = (-1,2)$, $\mathbf{c} = (-1,-1)$. Haz un dibujo.

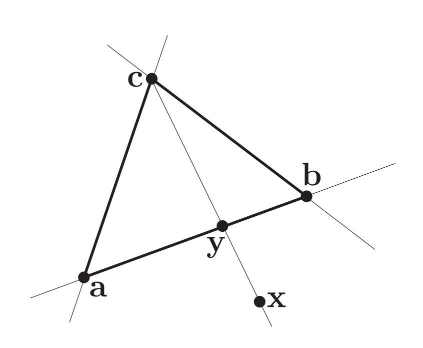
EJERCICIO 1.28 Observa que en el Teorema 1.5, así como en la discusión que le precede, nunca usamos que estuviéramos en el plano. ¿Podría el lector enunciar y demostrar el teorema análogo para tetraedros en el espacio? (Un tetraedro está determinado por cuatro puntos en el espacio tridimensional. ¿Dónde estará su centro de masa?) ¿Puede generalizarlo a más dimensiones?

1.4.2 Planos en el espacio I

En la demostración del Teorema de las Medianas (1.5) se usó una idea que podemos utilizar para definir planos en el espacio. Dados tres puntos $\mathbf{a}$, $\mathbf{b}$ y $\mathbf{c}$ en $\mathbb{R}^3$ (aunque debe observar el lector que nuestra discusión se generaliza a $\mathbb{R}^n$), ya sabemos describir las tres líneas entre ellos; supongamos que son distintas. Entonces podemos tomar nuevos puntos en alguna de ellas y de estos puntos las nuevas líneas que los unen al vértice restante (como lo hicimos del punto medio de un segmento al vértice opuesto en el teorema). Está claro que la unión de todas estas líneas debe ser el plano que pasa por $\mathbf{a}$, $\mathbf{b}$ y $\mathbf{c}$; ésta es la idea que vamos a desarrollar.



Un punto $\mathbf{y}$ en la línea por $\mathbf{a}$ y $\mathbf{b}$ se escribe como



$$\mathbf{y} = s\,\mathbf{a} + t\,\mathbf{b} \quad \text{con} \quad s + t = 1.$$

Y un punto $\mathbf{x}$ en la línea que pasa por $\mathbf{y}$ y $\mathbf{c}$ se escribe entonces como

$$\mathbf{x} = r\,(s\,\mathbf{a} + t\,\mathbf{b}) + (1-r)\,\mathbf{c},$$

para algún $r \in \mathbb{R}$; que es lo mismo que

$$\mathbf{x} = (r\,s)\,\mathbf{a} + (r\,t)\,\mathbf{b} + (1-r)\,\mathbf{c}.$$

Observemos que los tres coeficientes suman uno:

$$(r\,s) + (r\,t) + (1-r) = r\,(s + t) + 1 - r = r\,(1) + 1 - r = 1.$$

Y lo mismo hubiera pasado si en lugar de haber empezado con $\mathbf{a}$ y $\mathbf{b}$, hubiéramos comenzado con otra de las parejas. De lo que se trata entonces es que cualquier combinación de $\mathbf{a}$, $\mathbf{b}$ y $\mathbf{c}$ con coeficientes que sumen uno debe estar en su plano. Demostraremos que está en una línea por uno de los vértices y un punto en la línea que pasa por los otros dos.

Sean α, β, $\gamma \in \mathbb{R}$ tales que $\alpha + \beta + \gamma = 1$. Consideremos al punto

$$\mathbf{x} = \alpha\,\mathbf{a} + \beta\,\mathbf{b} + \gamma\mathbf{c}.$$

Como alguno de los coeficientes es distinto de 1 (si no, sumarían 3), podemos suponer sin pérdida de generalidad (esto quiere decir que los otros dos casos son análogos) que $\alpha \neq 1$. Entonces podemos dividir entre $1-\alpha$ y se tiene

$$\mathbf{x} = \alpha\,\mathbf{a} + (1-\alpha)\left(\frac{\beta}{1-\alpha}\,\mathbf{b} + \frac{\gamma}{1-\alpha}\mathbf{c}\right),$$

así que $\mathbf{x}$ está en la recta que pasa por $\mathbf{a}$ y el punto

$$\mathbf{y} = \frac{\beta}{1-\alpha}\,\mathbf{b} + \frac{\gamma}{1-\alpha}\mathbf{c}.$$

En este caso: $\alpha<0$, $\beta>0$, $\gamma>0$

Como $\alpha+\beta+\gamma = 1$, entonces $\beta+\gamma = 1-\alpha$, y los coeficientes de esta última expresión suman uno. Por lo tanto $\mathbf{y}$ está en la recta que pasa por $\mathbf{b}$ y $\mathbf{c}$. Hemos argumentado la siguiente definición:

Definición 1.4.2 Dados tres puntos $\mathbf{a}$, $\mathbf{b}$ y $\mathbf{c}$ en $\mathbb{R}^3$ *no colineales* (es decir, que no estén en una misma línea), el *plano* que pasa por ellos es el conjunto

$$\Pi = \{\alpha\,\mathbf{a} + \beta\,\mathbf{b} + \gamma\mathbf{c} \mid \alpha, \beta, \gamma \in \mathbb{R}\,;\ \alpha + \beta + \gamma = 1\}.$$

A una expresión de la forma $\alpha\,\mathbf{a} + \beta\,\mathbf{b} + \gamma\mathbf{c}$ con $\alpha + \beta + \gamma = 1$ la llamaremos *combinación afín* (o *baricéntrica*) de los puntos $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ y a los coeficientes α, β, γ las *coordenadas baricéntricas* del punto $\alpha\,\mathbf{a} + \beta\,\mathbf{b} + \gamma\mathbf{c}$.

EJERCICIO 1.29 Considera tres puntos $\mathbf{a}$, $\mathbf{b}$ y $\mathbf{c}$ no colineales, y sean α, β y γ las correspondientes coordenadas baricéntricas del plano que generan. Observa que cuando una de las coordenadas baricéntricas es cero, entonces el punto correspondiente está en una de las tres rectas por $\mathbf{a}$, $\mathbf{b}$ y $\mathbf{c}$. ¿En cuál?, ¿puedes demostrarlo? Haz un dibujo de los tres puntos y sus tres rectas. Éstas parten el plano en pedazos (¿cuántos?); en cada uno de ellos escribe los signos que toman las coordenadas baricéntricas (por ejemplo, en el interior del triángulo se tiene $(+,+,+)$ que corresponden respectivamente a (α, β, γ)).

EJERCICIO 1.30 Dibuja tres puntos $\mathbf{a}$, $\mathbf{b}$ y $\mathbf{c}$ no colineales. Con regla y compás encuentra los puntos $\mathbf{x} = (1/2)\mathbf{a} + (1/4)\,\mathbf{b} + (1/4)\,\mathbf{c}$, $\mathbf{y} = (1/4)\mathbf{a} + (1/2)\,\mathbf{b} + (1/4)\,\mathbf{c}$ y $\mathbf{z} = (1/4)\mathbf{a} + (1/4)\,\mathbf{b} + (1/2)\,\mathbf{c}$. Describe y argumenta tu construcción. Supón que puedes partir en tres el segmento $\overline{\mathbf{b}\mathbf{c}}$ (hazlo midiendo con una regla o a ojo). ¿Puedes construir los puntos $\mathbf{u} = (1/2)\mathbf{a} + (1/3)\,\mathbf{b} + (1/6)\,\mathbf{c}$ y $\mathbf{v} = (1/2)\mathbf{a} + (1/6)\,\mathbf{b} + (1/3)\,\mathbf{c}$?

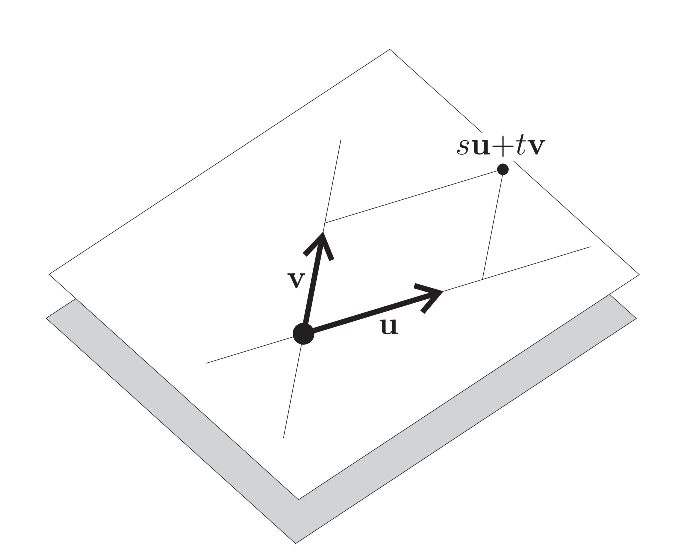
EJERCICIO 1.31 Demuestra que si $\mathbf{u}$ y $\mathbf{v}$ no son colineales con el origen, entonces el conjunto $\{s\,\mathbf{u} + t\,\mathbf{v} \mid s, t \in \mathbb{R}\}$ es un plano por el origen. (Usa que $s\,\mathbf{u} + t\,\mathbf{v} = r\,\mathbf{0} + s\,\mathbf{u} + t\,\mathbf{v}$ para cualquier r, s, $t \in \mathbb{R}$).

EJERCICIO 1.32 Demuestra que si $\mathbf{p}$ y $\mathbf{q}$ están en el plano Π de la Definición 1.4.2, entonces la recta que pasa por ellos está contenida en Π.

Definición lineal

Emocionado el autor por su demostración del teorema de medianas (1.5) usando coordenadas baricéntricas, se siguió de largo y definió planos en el espacio de una manera quizá no muy intuitiva. Que quede entonces la sección anterior como ejercicio en la manipulación de combinaciones de vectores; si no quedó muy clara puede releerse después de ésta. Pero conviene recapitular para definir los planos de otra manera, para aclarar la definición y así dejarnos la tarea de demostrar una equivalencia de dos definiciones.

La recta que genera un vector $\mathbf{u} \neq \mathbf{0}$ es el conjunto de todos sus "alargamientos" $\mathcal{L}_{\mathbf{u}} = \{s\,\mathbf{u} \mid s \in \mathbb{R}\}$. Si ahora tomamos un nuevo vector $\mathbf{v}$ que no esté en $\mathcal{L}_{\mathbf{u}}$, tenemos una nueva recta $\mathcal{L}_{\mathbf{v}} = \{t\,\mathbf{v} \mid t \in \mathbb{R}\}$ que intersecta a la anterior sólo en el origen. Estas dos rectas generan un plano que consiste en todos los puntos (en $\mathbb{R}^3$, digamos, para fijar ideas) a los cuales se puede llegar desde el origen moviéndose únicamente en las direcciones $\mathbf{u}$ y $\mathbf{v}$. Este plano que pasa por el origen claramente se describe con dos parámetros independientes:



$$\Pi_0 = \{s\,\mathbf{u} + t\,\mathbf{v} \mid s, t \in \mathbb{R}\},$$

que por su propia definición está hecho a imagen y semejanza del plano $\mathbb{R}^2$ (ver Ejercicio 1.33, adelante). Como lo hicimos con las rectas, podemos definir ahora un plano cualquiera como un plano por el origen (el que acabamos de describir) trasladado por cualquier otro vector constante.

Definición 1.4.3 Un *plano* en $\mathbb{R}^3$ es un conjunto de la forma

$$\Pi = \{\mathbf{p} + s\,\mathbf{u} + t\,\mathbf{v} \mid s, t \in \mathbb{R}\},$$

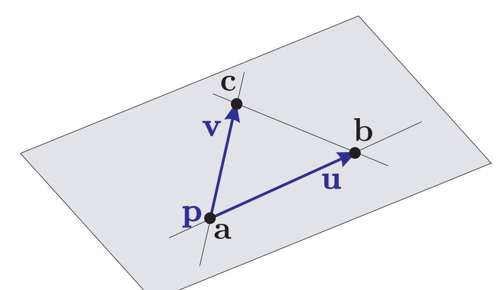
donde $\mathbf{u}$ y $\mathbf{v}$ son vectores no nulos tales que $\mathcal{L}_{\mathbf{u}} \cap \mathcal{L}_{\mathbf{v}} = \{\mathbf{0}\}$ y $\mathbf{p}$ es cualquier punto. A esta manera de describir un plano la llamaremos *expresión paramétrica*; a $\mathbf{u}$ y $\mathbf{v}$ se les llama *vectores direccionales* del plano Π y a $\mathbf{p}$ el *punto base* de la expresión paramétrica.

Para no confundirnos entre las dos definiciones de plano que hemos dado, llamemos *plano afín* a los que definimos en la sección anterior. Y ahora demostremos que las dos definiciones coinciden.

Lema 1.6 *En* $\mathbb{R}^3$*, todo plano es un plano afín y viceversa.*

Demostración. Sea Π como en la definición precedente. Debemos encontrar tres puntos en él y ver que todos los elementos de Π se expresan como combinación baricéntrica de ellos. Los puntos más obvios son

$$\mathbf{a} := \mathbf{p}\,;\ \mathbf{b} := \mathbf{p} + \mathbf{u}\,;\ \mathbf{c} := \mathbf{p} + \mathbf{v},$$



que claramente están en Π. Obsérvese que entonces se tiene que $\mathbf{u} = \mathbf{b} - \mathbf{a}$ y $\mathbf{v} = \mathbf{c} - \mathbf{a}$, de tal manera que para cualquier $s, t \in \mathbb{R}$ tenemos que

$$\begin{aligned} \mathbf{p} + s\,\mathbf{u} + t\,\mathbf{v} &= \mathbf{a} + s\,(\mathbf{b} - \mathbf{a}) + t\,(\mathbf{c} - \mathbf{a}) \\ &= \mathbf{a} + s\,\mathbf{b} - s\,\mathbf{a} + t\,\mathbf{c} - t\,\mathbf{a} \\ &= (1 - s - t)\,\mathbf{a} + s\,\mathbf{b} + t\,\mathbf{c}. \end{aligned}$$

Puesto que los coeficientes de esta última expresión suman 1, esto demuestra que Π está contenido en el plano afín generado por $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. E inversamente, cualquier combinación afín de $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ tiene esta última expresión, y por la misma igualdad se ve que está en Π. Obsérvese finalmente que si nos dan tres puntos $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, se pueden tomar como punto base a $\mathbf{a}$ y como vectores direccionales a $(\mathbf{b} - \mathbf{a})$ y $(\mathbf{c} - \mathbf{a})$ para expresar paramétricamente el plano afín que generan, pues las igualdades anteriores se siguen cumpliendo.

Ver que las condiciones que pusimos en ambas definiciones coinciden se deja como ejercicio. □

Antes de seguir adelante, conviene establecer cierta **terminología** y **notación** para cosas, nociones y expresiones que estamos usando mucho:

- Dados los vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ en $\mathbb{R}^n$ (para incluir nuestros casos de interés $n = 2, 3$ de una buena vez), a una expresión de la forma

  $$s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + s_k\mathbf{u}_k,$$

  donde $s_1, s_2, \ldots, s_k$ son números reales (escalares), se le llama una *combinación lineal* de los vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ con *coeficientes* $s_1, s_2, \ldots, s_k$. Obsérvese que toda combinación lineal da como resultado un vector, pero que un mismo vector tiene muchas expresiones tales.

- Como ya vimos, a una combinación lineal cuyos coeficientes suman 1 se le llama *combinación afín*. Y a una combinación afín de dos vectores distintos o de tres no colineales se le llama, además, *baricéntrica.*

- Al conjunto de *todas* las combinaciones lineales de los vectores $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ se le llama el *subespacio generado* por ellos y se le denotará $\langle \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k \rangle$. Es decir,

$$\langle \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k \rangle := \{ s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + s_k\mathbf{u}_k \mid s_1, s_2, \dots, s_k \in \mathbb{R} \} .$$

  Nótese que entonces, si $\mathbf{u} \neq \mathbf{0}$ se tiene que $\mathcal{L}_\mathbf{u} = \langle \mathbf{u} \rangle$ es la recta generada por $\mathbf{u}$; y ambas notaciones se seguirán usando indistintamente. Aunque ahora $\langle \mathbf{u} \rangle$ tiene sentido para $\mathbf{u} = \mathbf{0}$, en cuyo caso $\langle \mathbf{u} \rangle = \{\mathbf{0}\}$, y $\mathcal{L}_\mathbf{u}$ se usará para destacar que *es* una recta.

- Se dice que dos vectores $\mathbf{u}$ y $\mathbf{v}$ son *linealmente independientes* si son no nulos y tales que $\mathcal{L}_\mathbf{u} \cap \mathcal{L}_\mathbf{v} = \{\mathbf{0}\}$.

- Dado cualquier subconjunto $A \subset \mathbb{R}^n$, su *trasladado* por el vector (o al punto) $\mathbf{p}$ es el conjunto
$$A + \mathbf{p} = \mathbf{p} + A := \{ \mathbf{x} + \mathbf{p} \mid \mathbf{x} \in A \} .$$

Podemos resumir entonces nuestras dos definiciones básicas como: una *recta* es un conjunto de la forma $\ell = \mathbf{p} + \langle \mathbf{u} \rangle$ con $\mathbf{u} \neq \mathbf{0}$; y un *plano* es un conjunto de la forma $\Pi = \mathbf{p} + \langle \mathbf{u}, \mathbf{v} \rangle$ con $\mathbf{u}$ y $\mathbf{v}$ linealmente independientes.

EJERCICIO 1.33 Demuestra que si $\mathbf{u}$ y $\mathbf{v}$ son linealmente independientes, entonces la función $f : \mathbb{R}^2 \to \mathbb{R}^3$ definida por $f(s,t) = s\,\mathbf{u} + t\,\mathbf{v}$ es inyectiva.

EJERCICIO 1.34 Demuestra que cualquier plano está en biyección con $\mathbb{R}^2$.

EJERCICIO 1.35 Demuestra que tres puntos $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ son no colineales si y sólo si los vectores $\mathbf{u} := (\mathbf{b} - \mathbf{a})$ y $\mathbf{v} := (\mathbf{c} - \mathbf{a})$ son linealmente independientes.

EJERCICIO 1.36 Da una expresión paramétrica para el plano que pasa por los puntos $\mathbf{a} = (0,1,2)$, $\mathbf{b} = (1,1,0)$ y $\mathbf{c} = (-1,0,2)$.

EJERCICIO 1.37 Demuestra que $\mathbf{0} \in \langle \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k \rangle$ para cualquier $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$ en $\mathbb{R}^n$.

EJERCICIO 1.38 Demuestra que dos vectores $\mathbf{u}$ y $\mathbf{v}$ son linealmente independientes si y sólo si la única combinación lineal de ellos que da $\mathbf{0}$ es la *trivial* (*i.e.*, con ambos coeficientes cero).

EJERCICIO 1.39 Demuestra que

$$\mathbf{w} \in \langle \mathbf{u}, \mathbf{v} \rangle \quad \Leftrightarrow \quad \mathbf{w} + \langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle \quad \Leftrightarrow \quad \langle \mathbf{u}, \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle .$$

1.5 Medio Quinto

Regresemos al plano. Ya tenemos la noción de recta, y en esta sección veremos que nuestras rectas cumplen con la parte de existencia del quinto postulado de Euclides, que nuestra intuición va correspondiendo a nuestra formalización analítica de la geometría y que al cambiar los axiomas de Euclides por unos más básicos (los de los números reales) obtenemos los anteriores, pero ahora como teoremas demostrables. Ya vimos que por cualquier par de puntos se puede trazar un segmento que se extiende indefinidamente a ambos lados (axiomas I y III), es decir, que por ellos pasa una recta. El otro axioma que involucra rectas es el Quinto e incluye la noción de paralelismo, así que tendremos que empezar por ella.

Hay una definición conjuntista de rectas paralelas, asi que formalicémosla. Como las rectas son, por definición, ciertos subconjuntos distinguidos del plano, tiene sentido la siguiente:

Definición 1.5.1 Dos rectas ℓ_1 y ℓ_2 en $\mathbb{R}^2$ son *paralelas*, que escribiremos $\ell_1 \| \ell_2$, si no se intersectan, es decir si

$$\ell_1 \cap \ell_2 = \emptyset,$$

donde $\emptyset$ denota al conjunto vacío (ver Apéndice A).

Pero además de rectas tenemos algo más elemental que son los vectores (segmentos dirigidos) y entre ellos también hay una noción intuitiva de paralelismo que corresponde al "alargamiento" o multiplicación por escalares.

Definición 1.5.2 Dados dos vectores $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^2$ distintos de 0, diremos que $\mathbf{u}$ *es paralelo a* $\mathbf{v}$, lo que escribiremos $\mathbf{u} \| \mathbf{v}$, si existe un número $t \in \mathbb{R}$ tal que $\mathbf{u} = t\,\mathbf{v}$.

Hemos eliminado el vector 0, el origen, de la definición para no complicar la situación. Si lo hubiéramos incluido no es cierto el siguiente ejercicio.

EJERCICIO 1.40 Demuestra que la relación "ser paralelo a" es una relación de equivalencia en $\mathbb{R}^2 - \{0\}$ (en el plano menos el origen llamado el "plano agujerado"). Describe las clases de equivalencia. (En el Apéndice A se define relación de equivalencia.)

EJERCICIO 1.41 Sean $\mathbf{u}$,$\mathbf{v}$ dos vectores distintos de 0. Demuestra que:

$$\mathbf{u} \| \mathbf{v} \quad \Leftrightarrow \quad \mathcal{L}_{\mathbf{u}} \cap \mathcal{L}_{\mathbf{v}} \neq \{0\} \quad \Leftrightarrow \quad \mathcal{L}_{\mathbf{u}} = \mathcal{L}_{\mathbf{v}}.$$

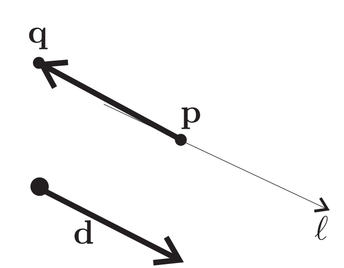
EJERCICIO 1.42 Demuestra que la relación entre rectas "ser paralelo a" es simétrica pero no reflexiva.

EJERCICIO 1.43 Demuestra que dos rectas *horizontales*, es decir, con vector direccional $\mathbf{d} = (1, 0)$, o son paralelas o son iguales.

Con la noción de paralelismo de vectores, podemos determinar cuándo un punto está en una recta.

Lema 1.7 *Sea ℓ la recta que pasa por $\mathbf{p}$ con dirección $\mathbf{d}$ ($\mathbf{d} \neq \mathbf{0}$), y sea $\mathbf{q} \neq \mathbf{p}$, entonces*

$$\mathbf{q} \in \ell \quad \Leftrightarrow \quad (\mathbf{q} - \mathbf{p}) \parallel \mathbf{d}.$$



Demostración. Tenemos que $\mathbf{q} \in \ell$ si y sólo si existe una $t \in \mathbb{R}$ tal que

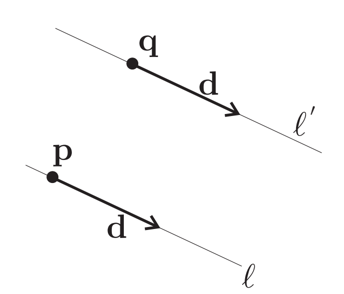
$$\mathbf{q} = \mathbf{p} + t\,\mathbf{d}\,.$$

Pero esto es equivalente a que $\mathbf{q} - \mathbf{p} = t\,\mathbf{d}$, que por definición es que $\mathbf{q} - \mathbf{p}$ es paralelo a $\mathbf{d}$. □

Teorema 1.8 (1/2 Quinto) *Sean ℓ una recta en $\mathbb{R}^2$ y $\mathbf{q}$ un punto fuera de ella, entonces existe una recta ℓ' que pasa por $\mathbf{q}$ y es paralela a ℓ.*

Demostración. Nuestra definición de recta nos da un punto $\mathbf{p}$ y un vector dirección $\mathbf{d} \neq \mathbf{0}$ para ℓ, de tal manera que

$$\ell = \{\mathbf{p} + t\,\mathbf{d} \mid t \in \mathbb{R}\}\,.$$



Es intuitivamente claro —y a la intuición hay que seguirla pues es, de cierta manera, lo que ya sabíamos— que la recta paralela deseada debe tener la misma dirección, así que definamos

$$\ell' = \{\mathbf{q} + s\,\mathbf{d} \mid s \in \mathbb{R}\}\,.$$

Como $\mathbf{q} \in \ell'$, nos falta demostrar que $\ell \| \ell'$, es decir, que $\ell \cap \ell' = \emptyset$. Para lograrlo, supongamos que no es cierto, es decir, que existe $\mathbf{x} \in \ell \cap \ell'$. Por las expresiones paramétricas de ℓ y ℓ', se tiene entonces que existen $t \in \mathbb{R}$ y $s \in \mathbb{R}$ para las cuales $\mathbf{x} = \mathbf{p} + t\,\mathbf{d}$ y $\mathbf{x} = \mathbf{q} + s\,\mathbf{d}$. De aquí se sigue que

$$\begin{aligned} \mathbf{q} + s\,\mathbf{d} &= \mathbf{p} + t\,\mathbf{d} \\ \mathbf{q} - \mathbf{p} &= t\,\mathbf{d} - s\,\mathbf{d} \\ \mathbf{q} - \mathbf{p} &= (t - s)\,\mathbf{d}, \end{aligned}$$

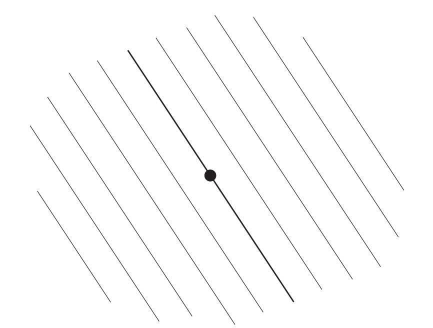
y entonces $\mathbf{q} \in \ell$ por el lema anterior, que es una contradicción a las hipótesis del teorema. Dicho de otra manera, como demostramos que $\ell \cap \ell' \neq \emptyset$ implica que $\mathbf{q} \in \ell$, podemos concluir que si $\mathbf{q} \notin \ell$, como en la hipótesis del teorema, no puede suceder que $\ell \cap \ell'$ sea no vacío, y por lo tanto $\ell \cap \ell' = \emptyset$ y $\ell \| \ell'$ por definición. Lo cual concluye con la parte de existencia del Quinto Postulado. □

La parte que falta demostrar del Quinto es la unicidad, es decir, que cualquier otra recta que pase por $\mathbf{q}$ sí intersecta a ℓ; que la única paralela es ℓ'. Esto se sigue de que rectas con vectores direccionales no paralelos siempre se intersectan, pero pospondremos la demostración formal de este hecho hasta tener más herramientas conceptuales. En particular, veremos en breve cómo encontrar la intersección de rectas para ejercitar la intuición, pero antes de entrarle, recapitulemos sobre la noción básica de esta sección, el paralelismo.

De la demostración del teorema se sigue que dos rectas con la misma dirección (o, lo que es lo mismo, con vectores direccionales paralelos) o son la misma o no se intersectan. Conviene entonces cambiar nuestra noción conjuntista de paralelismo a una vectorial.

Definición 1.5.3 Dos rectas ℓ_1 y ℓ_2 son *paralelas*, que escribiremos $\ell_1 \| \ell_2$, si tienen vectores direccionales paralelos.

Con esta nueva definición (que es la que se mantiene en lo sucesivo) una recta es paralela a sí misma, dos rectas paralelas distintas no se intersectan (son paralelas en el viejo sentido) y la relación es claramente transitiva. Así que es una relación de equivalencia y las clases de equivalencia corresponden a las clases de paralelismo de vectores (las rectas agujeradas por el origen). Llamaremos *haz de rectas paralelas* a una clase de paralelismo de rectas, es decir, al conjunto de todas las rectas paralelas a una dada (todas paralelas entre sí). Hay tantos haces de rectas paralelas como hay rectas por el origen, pues cada haz contiene exactamente una de estas rectas que, quitándole el origen, está formada por los posibles vectores direccionales para las rectas del haz.



Además, nuestra nueva noción de paralelismo tiene la gran ventaja de que se extiende a cualquier espacio vectorial. Nótese primero que la noción de paralelismo entre vectores no nulos se extiende sin problema a $\mathbb{R}^3$, pues está en términos del producto por escalares; y luego nuestra noción de paralelismo entre rectas se sigue de la de sus vectores direccionales. Por ejemplo, en el espacio $\mathbb{R}^3$ corresponde a nuestra noción intuitiva de paralelismo que no es conjuntista. Dos rectas pueden no intersectarse sin tener la misma dirección.

EJERCICIO 1.44 Da la descripción paramétrica de dos rectas en $\mathbb{R}^3$ que no se intersecten y que no sean paralelas.

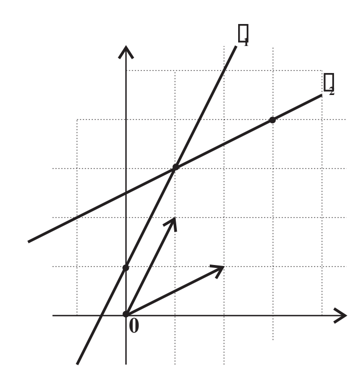
EJERCICIO 1.45 (**Quinto D3**) Demuestra que dados un plano Π en $\mathbb{R}^3$ y un punto $\mathbf{q}$ fuera de él, existe un plano Π' que pasa por $\mathbf{q}$ y que no intersecta a Π.

1.6 Intersección de rectas I

En esta sección se analiza el problema de encontrar la intersección de dos rectas. De nuevo, sabemos por experiencia e intuición que dos rectas no paralelas se deben intersectar en un punto. Cómo encontrar ese punto es la pregunta que responderemos en esta sección.



Hagamos un ejemplo. Sean



$$\begin{aligned} \ell_1 &= \{(0,1) + t(1,2) \mid t \in \mathbb{R}\} \\ \ell_2 &= \{(3,4) + s(2,1) \mid s \in \mathbb{R}\}. \end{aligned}$$

La intersección de estas dos rectas $\ell_1 \cap \ell_2$ es el punto que cumple con las dos descripciones. Es decir, buscamos una $t \in \mathbb{R}$ y una $s \in \mathbb{R}$ que satisfagan

$$(0,1) + t(1,2) = (3,4) + s(2,1),$$

donde es importante haber dado a los dos parámetros diferente nombre. Esta ecuación vectorial puede reescribirse como

$$\begin{aligned} t(1,2) - s(2,1) &= (3,4) - (0,1) \\ (t - 2s, 2t - s) &= (3,3), \end{aligned}$$

que nos da una ecuación lineal con dos incógnitas en cada coordenada. Es decir, debemos resolver el sistema:

$$\begin{aligned} t - 2s &= 3 \\ 2t - s &= 3. \end{aligned}$$

Si despejamos t en la primera ecuación, se obtiene $t = 3 + 2s$. Y al sustituir en la segunda obtenemos la ecuación lineal en la variable s:

$$2(3 + 2s) - s = 3,$$

que se puede resolver directamente:

$$\begin{aligned} 6 + 4s - s &= 3 \\ 3s &= -3 \\ s &= -1. \end{aligned}$$

Y por lo tanto, al sustituir este valor de s en la descripción de ℓ_2, el punto

$$(3,4) - (2,1) = (1,3)$$

está en las dos rectas. Nótese que bastó con encontrar el valor de un solo parámetro, pues de la correspondiente descripción paramétrica se obtiene el punto deseado. Pero debemos encontrar el mismo punto si resolvemos primero el otro parámetro. Veámoslo. Otra manera de resolver el sistema es eliminar la variable s. Para esto nos conviene multiplicar por -2 la segunda ecuación y sumarla a la primera, para obtener

$$-3t = -3,$$

de donde $t = 1$. Al sustituir en la parametrización de ℓ_1, nos da:

$$(0, 1) + (1, 2) = (1, 3),$$

como esperábamos.

Observemos primero que hay muchas maneras de resolver un sistema de dos ecuaciones lineales (se llaman así porque los exponentes de las incógnitas son uno —parece que no hay exponentes—) con dos incógnitas. Cualquiera de ellos es bueno y debe llevar a la misma solución.

Que en este problema particular aparezca un sistema de ecuaciones no es casualidad, pues en general encontrar la intersección de rectas se tiene que resolver así. Ya que si tenemos dos rectas cualesquiera en $\mathbb{R}^2$:

$$\begin{aligned} \ell_1 &= \{\mathbf{p} + t\,\mathbf{v} \mid t \in \mathbb{R}\} \\ \ell_2 &= \{\mathbf{q} + s\,\mathbf{u} \mid s \in \mathbb{R}\}. \end{aligned}$$

Su intersección está dada por la t y la s que cumplen

$$\mathbf{p} + t\,\mathbf{v} = \mathbf{q} + s\,\mathbf{u},$$

que equivale a

$$t\,\mathbf{v} - s\,\mathbf{u} = \mathbf{q} - \mathbf{p}. \tag{1.3}$$

Esta ecuación vectorial, a su vez, equivale a una ecuación lineal en cada coordenada. Como estamos en $\mathbb{R}^2$, y $\mathbf{p}$, $\mathbf{q}$, $\mathbf{v}$ y $\mathbf{u}$ son constantes, nos da entonces un sistema de dos ecuaciones lineales con las dos incógnitas t y s. En cada caso particular el sistema se puede resolver de diferentes maneras; por ejemplo, despejando una incógnita y sustituyendo, o bien tomando múltiplos de las ecuaciones y sumándolas para eliminar una variable y obtener una ecuación lineal con una sola variable (la otra) que se resuelve directamente.

Para hacer la teoría y entender qué pasa con estos sistemas en general, será bueno estudiar las ecuaciones lineales con dos incógnitas de manera geométrica; veremos que también están asociadas a las líneas rectas. Por el momento, de manera "algebraica", o mejor dicho, mecánica, resolvamos algunos ejemplos.

EJERCICIO 1.46 Encuentra los seis puntos de intersección de las cuatro rectas del Ejercicio 1.20.

EJERCICIO 1.47 Encuentra las intersecciones de las rectas $\ell_1 = \{(2,0) + t(1,-2) \mid t \in \mathbb{R}\}$, $\ell_2 = \{(2,1) + s(-2,4) \mid s \in \mathbb{R}\}$ y $\ell_3 = \{(1,2) + r(3,-6) \mid r \in \mathbb{R}\}$. Dibújalas para entender qué está pasando.

1.6.1 Sistemas de ecuaciones lineales I

Veamos ahora un método general para resolver sistemas.

Lema 1.9 *El sistema de ecuaciones*

$$\begin{aligned} a\,s + b\,t &= e \\ c\,s + d\,t &= f \end{aligned} \tag{1.4}$$

en las dos incógnitas s, t *tiene solución única si su* determinante $ad - bc$ *es distinto de cero.*

Demostración. Hay que intentar resolver el sistema general, seguir un método que funcione en todos los casos independientemente de los valores concretos que puedan tener las constantes a, b, c, d, e y f. Y el método más general es el de "eliminar" una variable para despejar la otra.

Para eliminar la t del sistema (1.4), multiplicamos por d la primera ecuación, y por $-b$ la segunda, para obtener

$$\begin{aligned} ad\,s + bd\,t &= ed \\ -bc\,s - bd\,t &= -bf, \end{aligned} \tag{1.5}$$

de tal manera que al sumar estas dos ecuaciones se tiene

$$(ad - bc)\,s = ed - bf\,. \tag{1.6}$$

Si $ad - bc \neq 0$,[3] entonces se puede despejar s:

$$s = \frac{ed - bf}{ad - bc}.$$

Y análogamente (¡hágalo como ejercicio!), o bien, sustituyendo el valor de s en cualquiera de las ecuaciones originales (¡convénzase!), se obtiene t:

$$t = \frac{af - ec}{ad - bc}\,.$$

Lo cual demuestra que si el determinante es distinto de cero, la solución es única; es precisamente la de las dos fórmulas anteriores. □

[3] Al número $ad - bc$ se le llama el *determinante* del sistema, porque determina que en este punto podemos proseguir.

Veamos ahora qué pasa con el sistema (1.4) cuando su determinante es cero. Si $ad - bc = 0$, obsérvese que aunque la ecuación (1.6) parezca muy sofisticada porque tiene muchas letras, en realidad es o una contradicción (algo falso) o una trivialidad. El coeficiente de s es 0, así que el lado izquierdo es 0. Entonces, si el lado derecho no es 0 es una contradicción; y si sí es 0, es cierto para cualquier s y hay una infinidad (precisamente un $\mathbb{R}$) de soluciones.

Lo que sucedió en este caso ($ad - bc = 0$) es que al intentar eliminar una de las variables también se eliminó la otra (como debió haberle sucedido al estudiante que hizo el Ejercicio 1.47). Entonces, o las dos ecuaciones son múltiplos y comparten soluciones (cuando $ed = bf$ las ecuaciones (1.5) ya son una la negativa de la otra); o bien, sólo los lados izquierdos son múltiplos, pero los derechos no lo son por el mismo factor ($ed \neq bf$) y no hay soluciones comunes.

Podemos resumir esto con el siguiente teorema que ya hemos demostrado.

Teorema 1.10 *Un sistema de dos ecuaciones lineales*

$$\begin{aligned} a\,s + b\,t &= e \\ c\,s + d\,t &= f \end{aligned} \tag{1.7}$$

en las dos incógnitas s, t *tiene solución única si y sólo si su determinante* $ad - bc$ *es distinto de cero. Además, si su determinante es cero (si* $ad - bc = 0$*) entonces no tiene solución o tiene una infinidad de ellas (tantas como* $\mathbb{R}$*).*

Si recordamos que los sistemas de dos ecuaciones lineales con dos incógnitas aparecieron al buscar la intersección de dos rectas, este teorema corresponde a nuestra intuición de las rectas: dos rectas se intersectan en un solo punto o no se intersectan o son la misma. De él dependerá la demostración de la mitad del Quinto que tenemos pendiente, pero conviene entrarle al toro por los cuernos: estudiar y entender primero el significado geométrico de las ecuaciones lineales con dos incógnitas.

EJERCICIO 1.48 Para las rectas de los dos ejercicios anteriores (1.46, 1.47), determina cómo se intersectan las rectas, usando únicamente el determinante.

1.7 Producto interior

En esta sección se introduce un ingrediente central para el estudio moderno de la geometría euclidiana: el "*producto interior*". Además de darnos la herramienta algebraica para el estudio geométrico de las ecuaciones lineales, en la sección siguiente nos dará mucho más. De él derivaremos después las nociones básicas de distancia y ángulo (de rigidez). Puede decirse que el producto interior es, aunque menos intuitivo, más elemental que las nociones de distancia y ángulo pues éstas se definirán en base a

aquél (aunque se verá también que estas dos juntas lo definen). El producto interior depende tan íntimamente de la idea cartesiana de coordenadas, que no tiene ningún análogo en la geometría griega. Pero tampoco viene de los primeros pininos que hizo la geometría analítica, pues no es sino hasta el siglo XIX cuando se le empezó a dar la importancia debida al desarrollarse las ideas involucradas en la noción general de espacio vectorial. Podría entonces decirse que el uso del producto interior (junto con el lenguaje de espacio vectorial) marca dos épocas en la geometría analítica. Pero es quizá la simpleza de su definición su mejor tarjeta de presentación.

Definición 1.7.1 Dados dos vectores $\mathbf{u} = (u_1, u_2)$ y $\mathbf{v} = (v_1, v_2)$, su *producto interno* (también conocido como producto *escalar* —que preferimos no usar para no confundir con la multiplicación por escalares— o bien como producto *punto*) es el número real:

$$\mathbf{u} \cdot \mathbf{v} = (u_1, u_2) \cdot (v_1, v_2) := u_1 v_1 + u_2 v_2.$$

En general, en $\mathbb{R}^n$ se define el *producto interior* (o el *producto punto*) de dos vectores $\mathbf{u} = (u_1, ..., u_n)$ y $\mathbf{v} = (v_1, ..., v_n)$ como la suma de los productos de sus coordenadas correspondientes, es decir,

$$\mathbf{u} \cdot \mathbf{v} := u_1 v_1 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i.$$

Así, por ejemplo:

$$(4, 3) \cdot (2, -1) = 4 \times 2 + 3 \times (-1) = 8 - 3 = 5,$$
$$(1, 2, 3) \cdot (4, 5, -6) = 4 + 10 - 18 = -4.$$

Obsérvese que el producto interior tiene como ingredientes dos vectores ($\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$) y nos da un escalar ($\mathbf{u} \cdot \mathbf{v} \in \mathbb{R}$); no debe confundirse con la multiplicación escalar, que de un escalar y un vector nos da un vector, excepto en el caso $n = 1$, en el que ambos coinciden con la multiplicación de los reales.

Antes de demostrar las propiedades básicas del producto interior, observemos que nos será muy útil para el problema que dejamos pendiente sobre sistemas de ecuaciones. En efecto en una ecuación lineal con dos incógnitas, x y y digamos, aparece una expresión de la forma $ax + by$ con a y b constantes. Si tomamos un vector constante $\mathbf{u} = (a, b)$ y un vector variable $\mathbf{x} = (x, y)$, ahora podemos escribir

$$\mathbf{u} \cdot \mathbf{x} = ax + by$$

y el “paquete básico” de información está en $\mathbf{u}$. En particular, como la mayoría de los lectores ya deben saber, la ecuación lineal

$$ax + by = c$$

donde c es una nueva constante, define una recta. Con el producto interior esta ecuación se reescribe como

$$\mathbf{u} \cdot \mathbf{x} = c.$$

Veremos cómo en el vector $\mathbf{u}$ y en la constante c se almacena la información geométrica de esa recta. Pero no nos apresuremos.

Como en el caso de la suma vectorial y la multiplicación por escalares, conviene demostrar primero las propiedades elementales del producto interior para manejarlo después con más soltura.

Teorema 1.11 *Para todos los vectores* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, *y para todo número* $t \in \mathbb{R}$ *se cumple que*

$$\begin{array}{rl} i) & \mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u} \\ ii) & \mathbf{u}\cdot(t\,\mathbf{v}) = t\,(\mathbf{u}\cdot\mathbf{v}) \\ iii) & \mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) = \mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w} \\ iv) & \mathbf{u}\cdot\mathbf{u} \geq 0 \\ v) & \mathbf{u}\cdot\mathbf{u} = 0 \Leftrightarrow \mathbf{u} = 0 \end{array}$$

Demostración. Nos interesa demostrarlo en $\mathbb{R}^2$ (y como ejercicio en $\mathbb{R}^3$), aunque el caso general es esencialmente lo mismo. Convendrá usar la notación de la misma letra con subíndices para las coordenadas, es decir, tomar $\mathbf{u} = (u_1, u_2)$, $\mathbf{v} = (v_1, v_2)$ y $\mathbf{w} = (w_1, w_2)$. El enunciado (**i**) se sigue inmediatamente de las definiciones y la conmutatividad de los números reales; (**ii**) se obtiene al factorizar:

$$\mathbf{u}\cdot(t\,\mathbf{v}) = u_1(tv_1) + u_2(tv_2) = t(u_1v_1 + u_2v_2) = t(\mathbf{u}\cdot\mathbf{v}).$$

De la distributividad y conmutatividad en los reales se obtiene (**iii**):

$$\begin{aligned} \mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) &= u_1(v_1+w_1) + u_2(v_2+w_2) \\ &= u_1v_1 + u_1w_1 + u_2v_2 + u_2w_2 \\ &= (u_1v_1 + u_2v_2) + (u_1w_1 + u_2w_2) = \mathbf{u}\cdot\mathbf{v} + \mathbf{u}\cdot\mathbf{w}. \end{aligned}$$

Al tomar el producto interior de un vector consigo mismo se obtiene

$$\mathbf{u}\cdot\mathbf{u} = u_1^2 + u_2^2,$$

que es una suma de cuadrados. Como cada cuadrado es positivo la suma también lo es (**iv**); y si fuera 0 significa que cada sumando es 0, es decir, que cada coordenada es 0 (**v**). □

EJERCICIO 1.49 Demuestra el Teorema 1.11 para $n = 3$.

EJERCICIO 1.50 Calcula el producto interior de algunos vectores del Ejercicio 1.3.

EJERCICIO 1.51 Demuestra sin usar (**v**), sólo la definición de producto interior, que dado $\mathbf{u} \in \mathbb{R}^2$ se tiene

$$\mathbf{u}\cdot\mathbf{x} = 0, \quad \forall \mathbf{x} \in \mathbb{R}^2 \Leftrightarrow \mathbf{u} = 0.$$

(Observa que sólo hay que usar el lado izquierdo para dos vectores muy sencillos.)

EJERCICIO 1.52 Demuestra el análogo del ejercicio anterior en $\mathbb{R}^3$. ¿Y en $\mathbb{R}^n$?

1.7.1 El compadre ortogonal

El primer uso geométrico que daremos al producto interior será para detectar la perpendicularidad. Es otra de las nociones básicas en los axiomas de Euclides que incluyen el concepto de ángulo recto.

Fijemos el vector $\mathbf{u} \in \mathbb{R}^2$, y para simplificar la notación digamos que $\mathbf{u} = (a, b) \neq \mathbf{0}$. Si tomamos $\mathbf{x} = (x, y)$ como un vector variable, vamos a ver que las soluciones de la ecuación

$$\mathbf{u} \cdot \mathbf{x} = 0, \tag{1.8}$$

es decir, los puntos en $\mathbb{R}^2$ cuyas coordenadas (x, y) satisfacen la ecuación

$$ax + by = 0,$$

forman una recta que es perpendicular a $\mathbf{u}$. Para esto consideraremos una solución particular (una de las más sencillas), $x = -b$, $y = a$; y será tan importante esta solución que le daremos nombre:

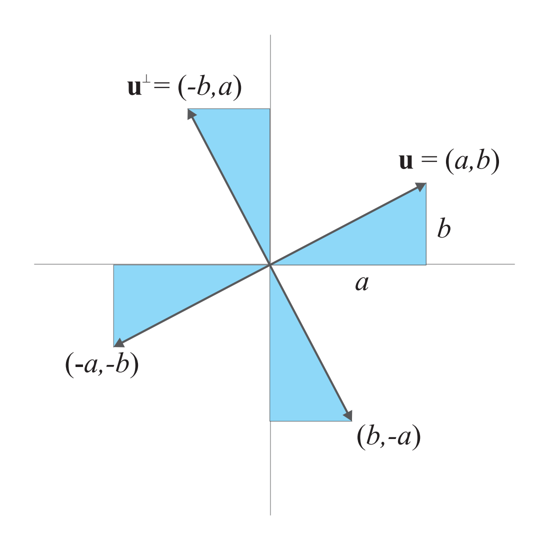
Definición 1.7.2 El *compadre ortogonal* del vector $\mathbf{u} = (a, b)$, denotado $\mathbf{u}^\perp$ y que se lee "$\mathbf{u}$–*perpendicular*" o "$\mathbf{u}$–*ortogonal*", es

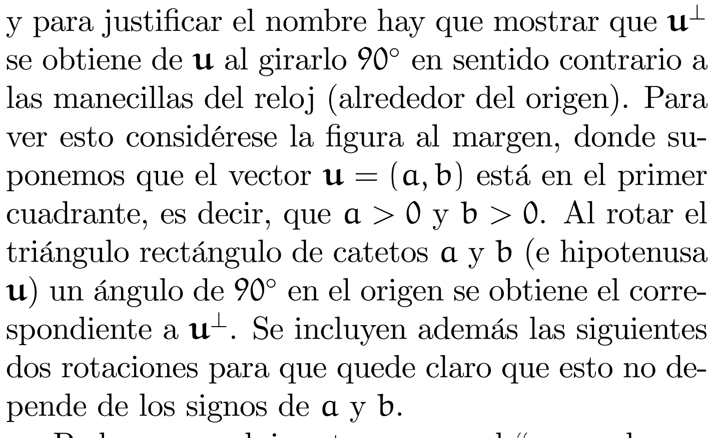
$$\mathbf{u}^\perp = (-b, a).$$

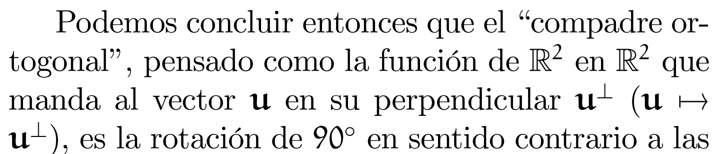
(Se intercambian coordenadas y a la primera se le cambia el signo.)

Se cumple que

$$\mathbf{u} \cdot \mathbf{u}^\perp = a(-b) + ba = 0;$$




y para justificar el nombre hay que mostrar que $\mathbf{u}^\perp$ se obtiene de $\mathbf{u}$ al girarlo $90°$ en sentido contrario a las manecillas del reloj (alrededor del origen). Para ver esto considérese la figura al margen, donde suponemos que el vector $\mathbf{u} = (a, b)$ está en el primer cuadrante, es decir, que $a > 0$ y $b > 0$. Al rotar el triángulo rectángulo de catetos a y b (e hipotenusa $\mathbf{u}$) un ángulo de $90°$ en el origen se obtiene el correspondiente a $\mathbf{u}^\perp$. Se incluyen además las siguientes dos rotaciones para que quede claro que esto no depende de los signos de a y b.


Podemos concluir entonces que el "compadre ortogonal", pensado como la función de $\mathbb{R}^2$ en $\mathbb{R}^2$ que manda al vector $\mathbf{u}$ en su perpendicular $\mathbf{u}^\perp$ ($\mathbf{u} \mapsto \mathbf{u}^\perp$), es la rotación de $90°$ en sentido contrario a las manecillas del reloj alrededor del origen. Es fácil comprobar que $(\mathbf{u}^\perp)^\perp = -\mathbf{u}$, ya sea con la fórmula o porque la rotación de $180°$ (rotar y volver a rotar $90°$) corresponde justo a tomar el inverso aditivo.

Como ya es costumbre, veamos algunas propiedades bonitas de la función "compadre ortogonal" respecto de las otras operaciones que hemos definido. La demostración, que consiste en dar coordenadas y aplicar definiciones, se deja al lector.

Lema 1.12 *Para todos los vectores* $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^2$, *y para todo número* $t \in \mathbb{R}$ *se cumple que*

$$\begin{array}{rl} \textit{i)} & (\mathbf{u}+\mathbf{v})^{\perp} = \mathbf{u}^{\perp} + \mathbf{v}^{\perp} \\ \textit{ii)} & (t\,\mathbf{u})^{\perp} = t\left(\mathbf{u}^{\perp}\right) \\ \textit{iii)} & \mathbf{u}^{\perp} \cdot \mathbf{v}^{\perp} = \mathbf{u} \cdot \mathbf{v} \\ \textit{iv)} & \mathbf{u}^{\perp} \cdot \mathbf{v} = -\left(\mathbf{u} \cdot \mathbf{v}^{\perp}\right). \end{array}$$

EJERCICIO 1.53 Demuestra el lema anterior.

Sistemas de ecuaciones lineales II

Usemos la herramienta del producto interior y el compadre ortogonal para revisar nuestro trabajo previo (sección 1.6.1) sobre sistemas de dos ecuaciones lineales con dos incógnitas. Si pensamos cada ecuación como la coordenada de un vector (y así fue como surgieron), un sistema tal se escribe

$$s\,\mathbf{u} + t\,\mathbf{v} = \mathbf{c}, \tag{1.9}$$

donde s y t son las incógnitas y $\mathbf{u}$, $\mathbf{v}$, $\mathbf{c} \in \mathbb{R}^2$ están dados. Para que este sistema sea justo el que ya estudiamos, (1.4), denotemos las coordenadas $\mathbf{u} = (a, c)$, $\mathbf{v} = (b, d)$; y para las constantes (con acrónimo $\mathbf{c}$), digamos que $\mathbf{c} = (e, f)$. Como ésta es una ecuación vectorial, si la multiplicamos (con el producto interior) por un vector, obtendremos una ecuación lineal real. Y para eliminar una variable, digamos s, podemos multiplicar por el compadre ortogonal de $\mathbf{u}$, para obtener

$$\begin{aligned} \mathbf{u}^{\perp} \cdot (s\,\mathbf{u}) + \mathbf{u}^{\perp} \cdot (t\,\mathbf{v}) &= \mathbf{u}^{\perp} \cdot \mathbf{c} \\ s\left(\mathbf{u}^{\perp} \cdot \mathbf{u}\right) + t\left(\mathbf{u}^{\perp} \cdot \mathbf{v}\right) &= \mathbf{u}^{\perp} \cdot \mathbf{c} \\ s\,(0) + t\left(\mathbf{u}^{\perp} \cdot \mathbf{v}\right) &= \mathbf{u}^{\perp} \cdot \mathbf{c} \\ t\left(\mathbf{u}^{\perp} \cdot \mathbf{v}\right) &= \mathbf{u}^{\perp} \cdot \mathbf{c}. \end{aligned}$$

Definición 1.7.3 El *determinante* de los vectores $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^2$ es el número real

$$\det(\mathbf{u}, \mathbf{v}) := \mathbf{u}^{\perp} \cdot \mathbf{v}.$$

Así que si suponemos que $\det(\mathbf{u}, \mathbf{v}) \neq 0$, podemos despejar t. De manera análoga podemos despejar s, y entonces concluir que el sistema de ecuaciones tiene solución única:

$$t = \frac{\mathbf{u}^{\perp} \cdot \mathbf{c}}{\mathbf{u}^{\perp} \cdot \mathbf{v}}, \; s = \frac{\mathbf{v}^{\perp} \cdot \mathbf{c}}{\mathbf{v}^{\perp} \cdot \mathbf{u}};$$

que es justo el Lema 1.9.

EJERCICIO 1.54 Escribe en coordenadas ($\mathbf{u} = (a, c)$, $\mathbf{v} = (b, d)$) cada uno de los pasos anteriores para ver que corresponden al método clásico de eliminación. (Quizá te convenga escribir las parejas como columnas, en vez de renglones, para que el sistema de ecuaciones sea claro.)

La ecuación lineal homogénea

Ahora sí, describamos el conjunto de soluciones de la ecuación lineal homogénea:

$$\mathbf{u} \cdot \mathbf{x} = 0.$$

Proposicion 1.13 *Sea* $\mathbf{u} \in \mathbb{R}^2 - \{\mathbf{0}\}$*, entonces*

$$\left\{\mathbf{x} \in \mathbb{R}^2 \mid \mathbf{u} \cdot \mathbf{x} = 0\right\} = \mathcal{L}_{\mathbf{u}^\perp},$$

donde, recuérdese, $\mathcal{L}_{\mathbf{u}^\perp} = \left\{t\,\mathbf{u}^\perp \mid t \in \mathbb{R}\right\}$.

Demostración. Tenemos que demostrar la igualdad de dos conjuntos. Esto se hace en dos pasos que corresponden a ver que cada elemento de un conjunto está también en el otro.

La contención más fácil es "$\supseteq$". Un elemento de $\mathcal{L}_{\mathbf{u}^\perp}$ es de la forma $t\,\mathbf{u}^\perp$ para algún $t \in \mathbb{R}$. Tenemos que demostrar que $t\,\mathbf{u}^\perp$ está en el conjunto de la izquierda. Para esto hay que ver que satisface la condición de pertenencia, que se sigue de **(ii)** en el Lema 1.12, y **(ii)** en el Teorema 1.11:

$$\mathbf{u} \cdot \left(t\,\mathbf{u}^\perp\right) = \mathbf{u} \cdot t\,\left(\mathbf{u}^\perp\right) = t\left(\mathbf{u} \cdot \mathbf{u}^\perp\right) = 0.$$

Para la otra contención, tomamos $\mathbf{x} \in \mathbb{R}^2$ tal que $\mathbf{u} \cdot \mathbf{x} = 0$ y debemos encontrar $t \in \mathbb{R}$ tal que $\mathbf{x} = t\,\mathbf{u}^\perp$. Sean $a, b \in \mathbb{R}$ las coordenadas de $\mathbf{u}$, es decir $\mathbf{u} = (a, b)$ y análogamente, sea $\mathbf{x} = (x, y)$. Entonces estamos suponiendo que

$$ax + by = 0.$$

Puesto que $\mathbf{u} \neq \mathbf{0}$ por hipótesis, tenemos que $a \neq 0$ o $b \neq 0$; y en cada caso podemos despejar una variable para encontrar el factor deseado:

$$a \neq 0 \Rightarrow x = -\frac{b\,y}{a},$$

y por lo tanto

$$\mathbf{x} = (x, y) = \frac{y}{a}\,(-b, a) = \left(\frac{y}{a}\right)\mathbf{u}^\perp;$$

o bien,

$$b \neq 0 \Rightarrow y = -\frac{a\,x}{b} \Rightarrow \mathbf{x} = (x, y) = -\frac{x}{b}\,(-b, a) = \left(-\frac{x}{b}\right)\mathbf{u}^\perp.$$

□

Hemos demostrado que la ecuación lineal *homogénea* (se llama así porque la constante es 0) $ax+by=0$, con a y b constantes reales, tiene como soluciones las parejas (x, y) que forman una recta por el origen de $\mathbb{R}^2$. Si empaquetamos la información de la ecuación en el vector $\mathbf{u}=(a, b)$, tenemos que $\mathbf{u} \neq \mathbf{0}$, pues de lo contrario no habría tal ecuación lineal (sería la tautología $0=0$ de lo que estamos hablando, y no es así). Pero además hemos visto que el vector $\mathbf{u}$ es ortogonal (tambien llamado *normal*) a la recta en cuestión, pues ésta está generada por el compadre ortogonal $\mathbf{u}^{\perp}$.

Se justifica entonces la siguiente definición, que era nuestro primer objetivo con el producto interior (donde, de una vez, estamos extrapolando a todas las dimensiones).

Definición 1.7.4 Se dice que dos vectores $\mathbf{u}$ y $\mathbf{v}$ en $\mathbb{R}^n$ son *perpendiculares* u *ortogonales*, y se escribe $\mathbf{u} \perp \mathbf{v}$, si $\mathbf{u} \cdot \mathbf{v}=0$.

Podemos reformular entonces la proposición anterior como "dada $\mathbf{u} \in \mathbb{R}^2$, $\mathbf{u} \neq \mathbf{0}$, el conjunto de $\mathbf{x} \in \mathbb{R}^2$ que satisfacen la ecuación $\mathbf{u} \cdot \mathbf{x}=0$ son la recta *perpendicular* a $\mathbf{u}$".

Antes de estudiar la ecuación general (no necesariamente homogénea), notemos que si el producto interior detecta perpendicularidad, también se puede usar, junto con el compadre ortogonal, para detectar paralelismo en el plano.

Corolario 1.14 *Sean* $\mathbf{u}$ *y* $\mathbf{v}$ *dos vectores no nulos en* $\mathbb{R}^2$*, entonces*

$$\mathbf{u} \parallel \mathbf{v} \quad \Leftrightarrow \quad \det(\mathbf{u}, \mathbf{v})=\mathbf{u}^{\perp} \cdot \mathbf{v}=0.$$

Demostración. Por la Proposición 1.13, $\mathbf{u}^{\perp} \cdot \mathbf{v}=0$ si y sólo si $\mathbf{v}$ pertenece a la recta generada por $\left(\mathbf{u}^{\perp}\right)^{\perp}$, que es la recta generada por $\mathbf{u}$ pues $\left(\mathbf{u}^{\perp}\right)^{\perp}=-\mathbf{u}$. □

Hay que hacer notar que si les damos coordenadas a $\mathbf{u}$ y a $\mathbf{v}$, digamos $\mathbf{u}=(a, b)$ y $\mathbf{v}=(c, d)$, entonces el determinante, que detecta paralelismo, es

$$\mathbf{u}^{\perp} \cdot \mathbf{v}=ad-bc,$$

que ya habíamos encontrado como determinante del sistema de ecuaciones (1.9), pero considerando $\mathbf{u}=(a, c)$ y $\mathbf{v}=(b, d)$.

EJERCICIO 1.55 Dibuja la recta definida como $\{\mathbf{x} \in \mathbb{R}^2 \mid \mathbf{u} \cdot \mathbf{x}=0\}$ para $\mathbf{u}$ con cada uno de los siguientes vectores: a) $(1,0)$, b) $(0,1)$, c) $(2,1)$, d) $(1,2)$, e) $(-1,1)$.

EJERCICIO 1.56 Describe el lugar geométrico definido como $\{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{u} \cdot \mathbf{x}=0\}$ para $\mathbf{u}$ con cada uno de los siguientes vectores: a) $(1,0,0)$, b) $(0,1,0)$, c) $(0,0,1)$, d) $(1,1,0)$.

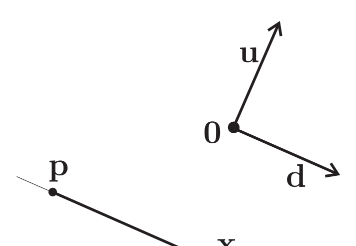
EJERCICIO 1.57 Sean $\mathbf{u}=(a, b)$ y $\mathbf{v}=(c, d)$ dos vectores no nulos. Sin usar el producto interior ni el compadre ortogonal —es decir, de la pura definición y con "álgebra elemental"— demuestra que $\mathbf{u}$ es paralelo a $\mathbf{v}$ si y sólo si $ad-bc=0$. Compara con la última parte de la demostración de la Proposición 1.13.

1.8 La ecuación normal de la recta

Demostraremos ahora que todas las rectas de $\mathbb{R}^2$ se pueden describir con una ecuación

$$\mathbf{u} \cdot \mathbf{x} = c,$$

donde $\mathbf{u} \in \mathbb{R}^2$ es un vector normal a la recta y la constante $c \in \mathbb{R}$ es escogida apropiadamente. Esta ecuación, vista en coordenadas, equivale a una ecuación lineal en dos variables ($ax + by = c$, al tomar $\mathbf{u} = (a, b)$ constante y $\mathbf{x} = (x, y)$ es el vector variable). Históricamente, el hecho de que las rectas tuvieran tal descripción fue una gran motivación en el inicio de la geometría analítica.



Tomemos una recta con dirección $\mathbf{d} \neq \mathbf{0}$

$$\ell = \{\mathbf{p} + t\,\mathbf{d} \mid t \in \mathbb{R}\}.$$

Si definimos $\mathbf{u} := \mathbf{d}^{\perp}$ y $c := \mathbf{u} \cdot \mathbf{p}$, afirmamos, es decir, demostraremos, que

$$\ell = \left\{\mathbf{x} \in \mathbb{R}^2 \mid \mathbf{u} \cdot \mathbf{x} = c\right\}.$$

Llamemos ℓ' a este último conjunto; demostrar la igualdad $\ell = \ell'$, equivale a demostrar las dos contenciones $\ell \subseteq \ell'$ y $\ell' \subseteq \ell$.

Si $\mathbf{x} = \mathbf{p} + t\,\mathbf{d} \in \ell$, entonces (usando las propiedades del producto interior y del compadre ortogonal) tenemos

$$\begin{aligned} \mathbf{u} \cdot \mathbf{x} &= \mathbf{u} \cdot (\mathbf{p} + t\,\mathbf{d}) = \mathbf{u} \cdot \mathbf{p} + t(\mathbf{u} \cdot \mathbf{d}) \\ &= \mathbf{u} \cdot \mathbf{p} + t(\mathbf{d}^{\perp} \cdot \mathbf{d}) = \mathbf{u} \cdot \mathbf{p} + 0 = c, \end{aligned}$$

lo cual demuestra que $\ell \subseteq \ell'$. Por el otro lado, dado $\mathbf{x} \in \ell'$ (*i.e.*, tal que $\mathbf{u} \cdot \mathbf{x} = c$), sea $\mathbf{y} := \mathbf{x} - \mathbf{p}$ (obsérvese que entonces $\mathbf{x} = \mathbf{p} + \mathbf{y}$). Se tiene que

$$\mathbf{u} \cdot \mathbf{y} = \mathbf{u} \cdot (\mathbf{x} - \mathbf{p}) = \mathbf{u} \cdot \mathbf{x} - \mathbf{u} \cdot \mathbf{p} = c - c = 0.$$

Por la Proposición 1.13, esto implica que $\mathbf{y}$ es paralelo a $\mathbf{u}^{\perp} = -\mathbf{d}$; y por tanto que $\mathbf{x} = \mathbf{p} + \mathbf{y} \in \ell$. Esto demuestra que $\ell = \ell'$.

En resumen, hemos demostrado que toda recta puede ser descrita por una ecuación normal:

Teorema 1.15 *Sea* $\ell = \{\mathbf{p} + t\,\mathbf{d} \mid t \in \mathbb{R}\}$ *una recta en* $\mathbb{R}^2$*. Entonces* ℓ *consta de los vectores* $\mathbf{x} \in \mathbb{R}^2$ *que cumplen la ecuación* $\mathbf{u} \cdot \mathbf{x} = c$*, que escribimos*

$$\ell : \quad \mathbf{u} \cdot \mathbf{x} = c,$$

donde $\mathbf{u} = \mathbf{d}^{\perp}$ *y* $c = \mathbf{u} \cdot \mathbf{p}$ *.* □

Este teorema también podría escribirse:

Teorema 1.16 *Sea* $\mathbf{d} \in \mathbb{R}^2 - \{0\}$ *, entonces*

$$\{\mathbf{p} + t\,\mathbf{d} \mid t \in \mathbb{R}\} = \left\{\mathbf{x} \in \mathbb{R}^2 \mid \mathbf{d}^\perp \cdot \mathbf{x} = \mathbf{d}^\perp \cdot \mathbf{p}\right\}$$

Estos enunciados nos dicen cómo encontrar una ecuación normal para una recta descrita paramétricamente. Por **ejemplo**, la recta

$$\{(s-2, 3-2s) \mid s \in \mathbb{R}\}$$

tiene como vector direccional al $(1, -2)$. Por tanto tiene vector normal a su compadre ortogonal $(2, 1)$, y como pasa por el punto $(-2, 3)$ entonces está determinada por la ecuación

$$\begin{aligned} (2,1)\cdot(x,y) &= (2,1)\cdot(-2,3) \\ 2x + y &= -1 \end{aligned}$$

(sustituya el lector las coordenadas de la descripción paramétrica en esta última ecuación). ...Dibujo

Para encontrar una representación paramétrica de la recta ℓ dada por una *ecuación normal*

$$\ell : \quad \mathbf{u} \cdot \mathbf{x} = c$$

(léase "ℓ dada por la ecuación $\mathbf{u} \cdot \mathbf{x} = c$"), bastará con encontrar una solución particular $\mathbf{p}$ (que es muy fácil porque se puede dar un valor arbitrario a una variable, 0 es la más conveniente, y despejar la otra), pues sabemos que la dirección es $\mathbf{u}^\perp$ (o cualquier paralelo). Por **ejemplo**, la recta dada por la ecuación normal

$$2x - 3y = 2$$

tiene vector normal $(2, -3)$. Por tanto tiene vector direccional $(3, 2)$; y como pasa por el punto $(1, 0)$, es el conjunto

$$\{(1 + 3t, 2t) \mid t \in \mathbb{R}\}.$$

Esto también se puede obtener directamente de las coordenadas, pero hay que partirlo en casos. Si $\mathbf{u} = (a, b)$, tenemos

$$\ell : \quad ax + by = c.$$

Supongamos que $a \neq 0$. Entonces se puede despejar x:

$$x = \frac{c}{a} - \frac{b}{a}y$$

y todas las soluciones de la ecuación se obtienen dando diferentes valores a y; es decir, son

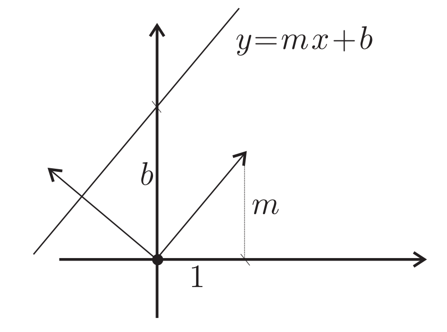
$$\left\{\left(\frac{c}{a} - \frac{b}{a}y, y\right) \mid y \in \mathbb{R}\right\} = \left\{\left(\frac{c}{a}, 0\right) + y\left(-\frac{b}{a}, 1\right) \mid y \in \mathbb{R}\right\},$$

que es una recta con dirección $(-b, a)$. Nos falta ver qué pasa cuando $a = 0$, pero entonces la recta es la horizontal $y = c/b$. Si hubieramos hecho el análisis cuando $b \neq 0$, se nos hubiera confundido la notación del caso clásico que vemos en el siguiente párrafo.

La ecuación "funcional" de la recta

Es bien conocida la ecuación *funcional de la recta*

$$y = mx + b,$$



que se usa para describir rectas como gráficas de una función; aquí m es la *pendiente* y b es la llamada "*ordenada al origen*" o el "*valor inicial*". En nuestros términos, esta ecuación se puede reescribir como

$$-mx + y = b$$

o bien, como

$$(-m, 1) \cdot (x, y) = b.$$

Por lo tanto, tiene vector normal $\mathbf{n} = (-m, 1)$. Podemos escoger a $\mathbf{d} = (1, m) = -\mathbf{n}^{\perp}$ como su vector direccional y a $\mathbf{p} = (0, b)$ como una solución particular. Así que su parametrización natural es (usando a x como parámetro):

$$\{(0, b) + x(1, m) \mid x \in \mathbb{R}\} = \{(x, mx + b) \mid x \in \mathbb{R}\},$$

que es la gráfica de una función. Obsérvese que todas las rectas excepto las verticales (con dirección $(0, 1)$) se pueden expresar así.

No está de más observar que para cualquier función $f : \mathbb{R} \to \mathbb{R}$ se obtiene un subconjunto de $\mathbb{R}^2$, llamado la *gráfica de la función* f definida paramétricamente como

$$\{(x, f(x)) \mid x \in \mathbb{R}\};$$

y entonces hemos visto que todas las rectas, excepto las verticales, son la gráfica de funciones de la forma $f(x) = mx+b$, que en el capítulo 3 llamaremos funciones afines.

EJERCICIO 1.58 Para las siguientes rectas, encuentra una ecuación normal y, en su caso, su ecuación funcional:

a) $\{(2, 3) + t(1, 1) \mid t \in \mathbb{R}\}$
b) $\{(-1, 0) + s(2, 1) \mid s \in \mathbb{R}\}$
c) $\{(0, -2) + (-r, 2r) \mid r \in \mathbb{R}\}$
d) $\{(1, 3) + s(2, 0) \mid s \in \mathbb{R}\}$
e) $\{(t - 1, -2t) \mid t \in \mathbb{R}\}$

EJERCICIO 1.59 Da una descripción paramétrica de las rectas dadas por las ecuaciones:

a) $2x - 3y = 1$
b) $2x - y = 2$
c) $2y - 4x = 2$
d) $x + 5y = -1$
e) $3 - 4y = 2x + 2$

EJERCICIO 1.60 Encuentra una ecuación normal para la recta que pasa por los puntos:
a) $(2,1)$ y $(3,4)$
b) $(-1,1)$ y $(2,2)$
c) $(1,-3)$ y $(3,1)$
d) $(2,0)$ y $(1,1)$

EJERCICIO 1.61 Sean $\mathbf{p}$ y $\mathbf{q}$ dos puntos distintos en $\mathbb{R}^2$. Demuestra que la recta ℓ que pasa por ellos tiene ecuación normal:

$$\ell: \quad (\mathbf{q}-\mathbf{p})^\perp \cdot \mathbf{x} = \mathbf{q}^\perp \cdot \mathbf{p}.$$

EJERCICIO 1.62 Encuentra la intersección de las rectas (a) y (b) del Ejercicio 1.59.

EJERCICIO 1.63 Encuentra la intersección de la recta (α) del Ejercicio 1.58 con la de la recta (α) del Ejercicio 1.59, donde $\alpha \in \{a,b,c,d\}$.

EJERCICIO 1.64 Dibuja las gráficas de las funciones $f(x) = x^2$ y $g(x) = x^3$.

1.8.1 Intersección de rectas II

Regresemos ahora al problema teórico de encontrar la intersección de rectas, pero con la nueva herramienta de la ecuación normal. Consideremos dos rectas ℓ_1 y ℓ_2 en $\mathbb{R}^2$. Sean $\mathbf{u} = (a,b)$ y $\mathbf{v} = (c,d)$ vectores normales a ellas respectivamente, de tal manera que ambos son no nulos y existen constantes $e, f \in \mathbb{R}$ para las cuales

$$\begin{aligned} \ell_1 &: \quad \mathbf{u} \cdot \mathbf{x} = e \\ \ell_2 &: \quad \mathbf{v} \cdot \mathbf{x} = f. \end{aligned}$$

Es claro entonces que $\ell_1 \cap \ell_2$ consta de los puntos $\mathbf{x} \in \mathbb{R}^2$ que satisfacen ambas ecuaciones. Si llamamos, como de costumbre, x y y a las coordenadas de $\mathbf{x}$ (es decir, si hacemos $\mathbf{x} = (x,y)$), las dos ecuaciones anteriores son el sistema

$$\begin{aligned} ax + by &= e \\ cx + dy &= f, \end{aligned}$$

con a, b, c, d, e, f constantes y x, y variables o incógnitas. Éste es precisamente el sistema que estudiamos en la sección 1.6.1; aunque allá hablabamos de los "parámetros s y t", son esencialmente lo mismo. Pero ahora tiene el significado geométrico que buscábamos: resolverlo es encontrar el punto de intersección de dos rectas (sus coordenadas tal cual, y no los parámetros para encontrarlo, como antes). Como ya hicimos el trabajo abstracto de resolver el sistema, sólo nos queda por hacer la traducción al lenguaje geométrico. Como ya vimos, el tipo de solución (una única, ninguna o tantas como reales) depende del determinante del sistema $ad-bc$, y éste, a su vez, ya se nos apareció (nótese que $\det(\mathbf{u},\mathbf{v}) = \mathbf{u}^\perp \cdot \mathbf{v} = ad-bc$) como el "numero mágico" que detecta paralelismo (Corolario 1.14). De tal manera que del Corolario 1.14 y del Teorema 1.10 podemos concluir:

Teorema 1.17 *Dadas las rectas*

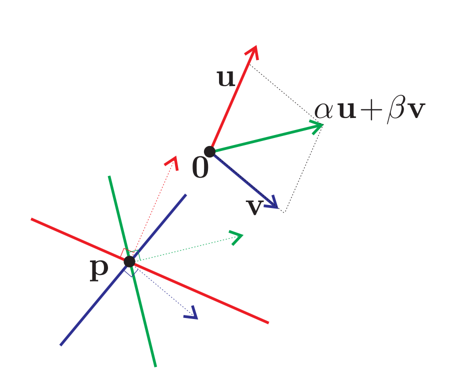
$$\begin{array}{lcl} \ell_1 & : & \mathbf{u} \cdot \mathbf{x} = e \\ \ell_2 & : & \mathbf{v} \cdot \mathbf{x} = f, \end{array}$$

entonces:

i) $\det(\mathbf{u}, \mathbf{v}) \neq 0 \;\Leftrightarrow\; \ell_1 \cap \ell_2$ *es un único punto;*
ii) $\det(\mathbf{u}, \mathbf{v}) = 0 \;\Leftrightarrow\; \ell_1 \parallel \ell_2 \Leftrightarrow\; \ell_1 \cap \ell_2 = \emptyset$ *o* $\ell_1 = \ell_2$.

Consideremos con más detenimiento el segundo caso, cuando el determinante $\mathbf{u}^{\perp} \cdot \mathbf{v}$ es cero y por tanto los vectores normales (y las rectas) son paralelos. Tenemos entonces que para alguna $t \in \mathbb{R}$, se cumple que $\mathbf{v} = t\,\mathbf{u}$ (nótese que $t \neq 0$, pues ambos vectores son no nulos). La disyuntiva de si las rectas no se intersectan o son iguales corresponde a que las constantes e y f difieran por el mismo factor, es decir si $f \neq t\,e$ o $f = t\,e$ respectivamente. Podemos sacar dos conclusiones interesantes. Primero, dos ecuaciones normales definen la misma recta si y sólo si las tres constantes que las determinan (en nuestro caso (a, b, e) y (c, d, f)) son vectores paralelos en $\mathbb{R}^3$. Y segundo, que al fijar un vector $\mathbf{u} = (a, b)$ y variar un parámetro $c \in \mathbb{R}$, las ecuaciones $\mathbf{u} \cdot \mathbf{x} = c$ determinan el haz de rectas paralelas que es ortogonal a $\mathbf{u}$.

Otro punto interesante que vale la pena discutir es el método que usamos para resolver sistemas de ecuaciones. Se consideraron ciertos múltiplos de ellas y luego se sumaron. Es decir, para ciertas α y β en $\mathbb{R}$ se obtiene, de las dos ecuaciones dadas, una nueva de la forma



$$(\alpha\,\mathbf{u} + \beta\,\mathbf{v}) \cdot \mathbf{x} = \alpha\,e + \beta\,f.$$

Ésta es la ecuación de otra recta. La propiedad importante que cumple es que si $\mathbf{p}$ satisface las dos ecuaciones dadas, entonces también satisface esta última por las propiedades de nuestras operaciones. De tal manera que esta nueva ecuación define una recta que pasa por el punto de intersección $\mathbf{p}$. El método de eliminar una variable consiste en encontrar la horizontal o la vertical que pasa por el punto (la mitad del problema). Pero en general, todas las ecuaciones posibles de la forma anterior definen el *haz de rectas concurrentes* por el punto de intersección $\mathbf{p}$ cuando $\mathbf{u}$ y $\mathbf{v}$ no son paralelos, pues los vectores $\alpha\,\mathbf{u} + \beta\,\mathbf{v}$ (al variar α y β) son más que suficientes para definir todas las posibles direcciones.

Por último, y aunque ya hayamos demostrado la parte de existencia, concluyamos con el Quinto usando al Teorema 1.17.

Teorema 1.18 *Dada una recta* ℓ *y un punto* $\mathbf{p}$ *fuera de ella en el plano, existe una única recta* ℓ' *que pasa por* $\mathbf{p}$ *y no intersecta a* ℓ*.*

Demostración. Existe un vector $\mathbf{u} \neq \mathbf{0}$ y una constante $c \in \mathbb{R}$, tales que ℓ está dada por la ecuación $\mathbf{u} \cdot \mathbf{x} = c$. Sea

$$\ell' : \quad \mathbf{u} \cdot \mathbf{x} = \mathbf{u} \cdot \mathbf{p}.$$

Entonces $\mathbf{p} \in \ell'$ porque satisface la ecuación, $\ell \cap \ell' = \emptyset$ porque $\mathbf{u} \cdot \mathbf{p} \neq c$ (pues $\mathbf{p} \notin \ell$), y cualquier otra recta que pasa por $\mathbf{p}$ intersecta a ℓ pues su vector normal no es paralelo a $\mathbf{u}$. □

EJERCICIO 1.65 Encuentra la intersección de las rectas (a), (b), (c) y (d) del Ejercicio 1.59.

EJERCICIO 1.66 Demuestra que dos vectores $\mathbf{u}$ y $\mathbf{v}$ en $\mathbb{R}^2$ son linealmente independientes si y solo si $\mathbf{u}^\perp \cdot \mathbf{v} \neq 0$.

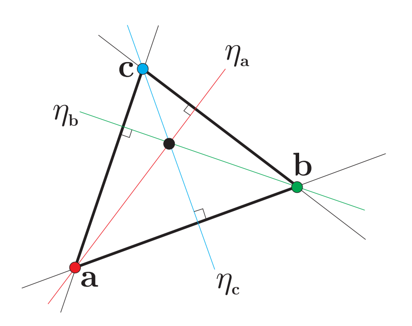
EJERCICIO 1.67 Demuestra que dos vectores $\mathbf{u}$ y $\mathbf{v}$ en $\mathbb{R}^2$ son linealmente independientes si y sólo si $\mathbb{R}^2 = \langle \mathbf{u}, \mathbf{v} \rangle$.

EJERCICIO 1.68 A cada terna de números reales (a, b, c), asóciale la ecuación $a\,x + b\,y = c$.
i) ¿Cuáles son las ternas (como puntos en $\mathbb{R}^3$) a las que se le asocian rectas en $\mathbb{R}^2$ por su ecuación normal?
ii) Dada una recta en $\mathbb{R}^2$, describe las ternas (como subconjunto de $\mathbb{R}^3$) que se asocian a ella.
iii) Dado un haz de rectas paralelas en $\mathbb{R}^2$, describe las ternas (como subconjunto de $\mathbb{R}^3$) que se asocian a rectas de ese haz.
iv) Dado un haz de rectas concurrentes en $\mathbb{R}^2$, describe las ternas (como subconjunto de $\mathbb{R}^3$) que se asocian a las rectas de ese haz.

1.8.2 Teoremas de concurrencia

En esta sección, aplicamos la ecuación normal de las rectas para demostrar uno de los teoremas clásicos de concurrencia de líneas y dejamos el otro como ejercicio.

La *altura* de un triángulo es la recta que pasa por uno de sus vértices y es ortogonal al lado opuesto.



Teorema 1.19 *Las alturas de un triángulo son concurrentes.*

Demostración. Dado un triángulo con vértices $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, tenemos que la altura por el vértice $\mathbf{a}$, llamémosla η_a ("eta-sub-a"), está definida por

$$\eta_a : \quad (\mathbf{c} - \mathbf{b}) \cdot \mathbf{x} = (\mathbf{c} - \mathbf{b}) \cdot \mathbf{a},$$

pues es ortogonal al lado que pasa por $\mathbf{b}$ y $\mathbf{c}$, que tiene dirección $(\mathbf{c}-\mathbf{b})$, y pasa por el punto $\mathbf{a}$. Analogamente:

$$\begin{array}{rcl} \eta_{\mathbf{b}} & : & (\mathbf{a}-\mathbf{c})\cdot\mathbf{x}=(\mathbf{a}-\mathbf{c})\cdot\mathbf{b} \\ \eta_{\mathbf{c}} & : & (\mathbf{b}-\mathbf{a})\cdot\mathbf{x}=(\mathbf{b}-\mathbf{a})\cdot\mathbf{c}. \end{array}$$

Obsérvese ahora que la suma (lado a lado) de dos de estas ecuaciones da precisamente el negativo de la tercera. Por ejemplo, sumando las dos primeras obtenemos

$$\begin{aligned} (\mathbf{c}-\mathbf{b})\cdot\mathbf{x}+(\mathbf{a}-\mathbf{c})\cdot\mathbf{x} &= (\mathbf{c}-\mathbf{b})\cdot\mathbf{a}+(\mathbf{a}-\mathbf{c})\cdot\mathbf{b} \\ (\mathbf{c}-\mathbf{b}+\mathbf{a}-\mathbf{c})\cdot\mathbf{x} &= \mathbf{c}\cdot\mathbf{a}-\mathbf{b}\cdot\mathbf{a}+\mathbf{a}\cdot\mathbf{b}-\mathbf{c}\cdot\mathbf{b} \\ (\mathbf{a}-\mathbf{b})\cdot\mathbf{x} &= (\mathbf{a}-\mathbf{b})\cdot\mathbf{c}. \end{aligned}$$

Así que si $\mathbf{x}\in\eta_{\mathbf{a}}\cap\eta_{\mathbf{b}}$, entonces cumple las dos primeras ecuaciones y por tanto cumple su suma que es "menos" la ecuación de $\eta_{\mathbf{c}}$, y entonces $\mathbf{x}\in\eta_{\mathbf{c}}$. Así que las tres rectas pasan por el mismo punto. □

EJERCICIO 1.69 Demuestra que las tres mediatrices de un triángulo son concurrentes (el punto en el que concurren se llama *circuncentro*). Donde la mediatriz de un segmento es su ortogonal que pasa por el punto medio.

EJERCICIO 1.70 Encuentra el circuncentro del triángulo con vértices $(1,1)$, $(1,-1)$ y $(-2,-2)$. Haz el dibujo del triángulo y sus mediatrices.

1.8.3 Planos en el espacio II

Hemos definido las rectas en $\mathbb{R}^n$ por su representación paramétrica y, en $\mathbb{R}^2$, acabamos de ver que también tienen una representación normal (con base en una ecuación lineal). Es entonces importante hacer notar que este fenómeno sólo se da en dimensión 2. En $\mathbb{R}^3$, que es el otro espacio que nos interesa, los planos (¡no las rectas!) son las que se definen por la ecuación normal. Veamos esto con cuidado.

Dado un vector $\mathbf{n}\in\mathbb{R}^3$ distinto de $\mathbf{0}$; sea, para cualquier $d\in\mathbb{R}$,

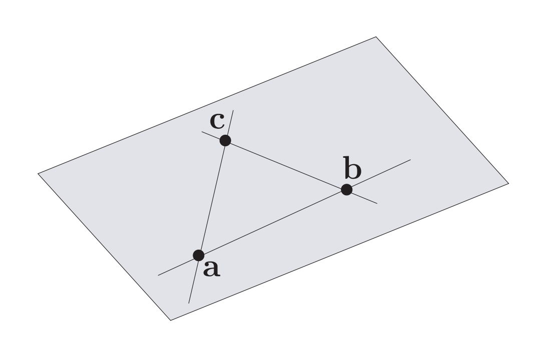
$$\Pi_d: \quad \mathbf{n}\cdot\mathbf{x}=d$$

es decir, $\Pi_d := \{\mathbf{x}\in\mathbb{R}^3 \mid \mathbf{n}\cdot\mathbf{x}=d\}$. Llamamos a la constante d pues ahora el vector normal es de la forma $\mathbf{n}=(a,b,c)$ y entonces la ecuación anterior se escribe en coordenadas como

$$ax+by+cz=d$$

con el vector variable $\mathbf{x}=(x,y,z)$. Afirmamos que Π_d es un plano. Recuérdese que definimos plano como el conjunto de combinaciones afines (o baricéntricas) de tres puntos no colineales.

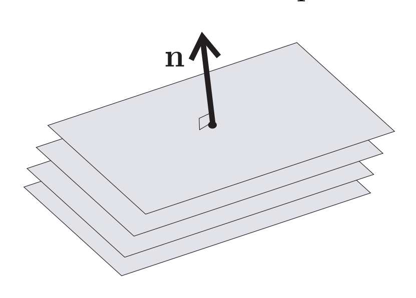
Veamos primero que si tres puntos $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ satisfacen la ecuación $\mathbf{n} \cdot \mathbf{x} = d$, entonces los puntos del plano que generan también la satisfacen. Para esto, tomamos $\alpha, \beta, \gamma \in \mathbb{R}$ tales que $\alpha + \beta + \gamma = 1$, y entonces se tiene



$$\begin{aligned} \mathbf{n} \cdot (\alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}) &= \alpha(\mathbf{n} \cdot \mathbf{a}) + \beta(\mathbf{n} \cdot \mathbf{b}) + \gamma(\mathbf{n} \cdot \mathbf{c}) \\ &= \alpha d + \beta d + \gamma d \\ &= (\alpha + \beta + \gamma)\ d = d. \end{aligned}$$

Esto demuestra que si $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \Pi_d$ entonces el plano (o si son colineales, la recta) que generan está contenido en Π_d. Faltaría ver que en Π_d hay tres puntos no colineales y luego demostrar que no hay más soluciones que las del plano que generan.

Pero mejor veámoslo desde otro punto de vista: el lineal en vez del baricéntrico. Afirmamos que los conjuntos Π_d, al variar d, son la familia de planos normales a $\mathbf{n}$. Para demostrar esto, hay que ver que Π_0 es un plano por el origen y que Π_d es un trasladado de Π_0, es decir, Π_0 empujado por algún vector constante.



Esto último ya se hizo en esencia cuando vimos el caso de rectas en el plano, así que lo veremos primero para remarcar lo general de aquella demostración. Supongamos que $\mathbf{p} \in \mathbb{R}^3$ es tal que

$$\mathbf{n} \cdot \mathbf{p} = d,$$

es decir, que es una solución particular. Vamos a demostrar que

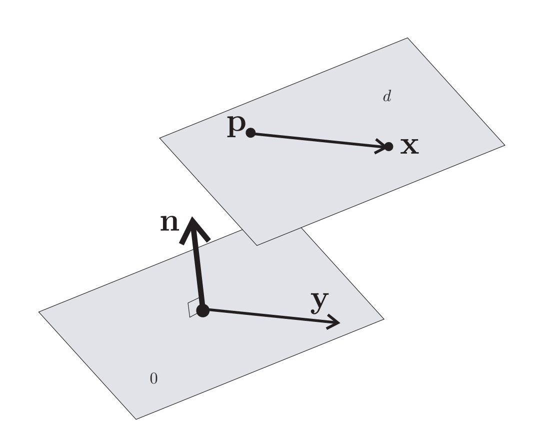
$$\Pi_d = \Pi_0 + \mathbf{p} := \{\mathbf{y} + \mathbf{p} \mid \mathbf{y} \in \Pi_0\}; \tag{1.10}$$

es decir, que si $\mathbf{x} \in \Pi_d$ entonces $\mathbf{x} = \mathbf{y} + \mathbf{p}$ para algún $\mathbf{y} \in \Pi_0$; y al revés, que si $\mathbf{y} \in \Pi_0$ entonces $\mathbf{x} = \mathbf{y} + \mathbf{p} \in \Pi_d$.

Dado $\mathbf{x} \in \Pi_d$ (*i.e.*, tal que $\mathbf{n} \cdot \mathbf{x} = d$), sea $\mathbf{y} := \mathbf{x} - \mathbf{p}$. Como

$$\mathbf{n} \cdot \mathbf{y} = \mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = \mathbf{n} \cdot \mathbf{x} - \mathbf{n} \cdot \mathbf{p} = d - d = 0,$$

entonces $\mathbf{y} \in \Pi_0$ y es, por definición, tal que $\mathbf{x} = \mathbf{y} + \mathbf{p}$; por lo tanto $\Pi_d \subseteq \Pi_0 + \mathbf{p}$. Y al revés, si $\mathbf{y} \in \Pi_0$ y $\mathbf{x} = \mathbf{y} + \mathbf{p}$, entonces



$$\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot (\mathbf{y} + \mathbf{p}) = \mathbf{n} \cdot \mathbf{y} + \mathbf{n} \cdot \mathbf{p} = 0 + d = d$$

y por lo tanto $\mathbf{x} \in \Pi_d$. Hemos demostrado (1.10).

Nos falta ver que Π_0 es efectivamente un plano por el origen. Puesto que estamos suponiendo que $\mathbf{n} \neq \mathbf{0}$ entonces de la ecuación

$$ax + by + cz = 0$$

puede despejarse alguna variable en términos de las otras $\mathtt{dos}$; sin pérdida de generalidad, digamos que es z, es decir, que $c \neq 0$. De tal forma que al dar valores arbitrarios a x y y, la fórmula nos da un valor de z, y por tanto un punto $\mathbf{x} \in \mathbb{R}^3$ que satisface la condición original $\mathbf{n} \cdot \mathbf{x} = 0$ —esto equivale a parametrizar las soluciones con $\mathbb{R}^2$, con dos grados de libertad—. En particular hay una solución $\mathbf{u}$ con $x = 1$ y $y = 0$ (a saber, $\mathbf{u} = (1, 0, -a/c)$) y una solución $\mathbf{v}$ con $x = 0$ y $y = 1$ (¿cuál es?). Como claramente los vectores $\mathbf{u}$ y $\mathbf{v}$ no son paralelos (tienen ceros en coordenadas distintas), generan linealmente todo un plano de soluciones (pues para cualquier $s, t \in \mathbb{R}$, se tiene que $\mathbf{n} \cdot (s\,\mathbf{u} + t\,\mathbf{v}) = s\,(\mathbf{n} \cdot \mathbf{u}) + t\ (\mathbf{n} \cdot \mathbf{v}) = 0$). Falta ver que no hay más soluciones que las que hemos descrito, pero esto se lo dejamos al estudiante para el siguiente ejercicio, que hay que comparar con este párrafo para entender cómo se llegó a su elegante planteamiento.

EJERCICIO 1.71 Sea $\mathbf{n} = (a, b, c)$ tal que $a \neq 0$. Demuestra que

$$\left\{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{n} \cdot \mathbf{x} = 0\right\} = \{\ s\ (-b, a, 0) + t\ (-c, 0, a) \mid s, t \in \mathbb{R}\}.$$

EJERCICIO 1.72 Describe el plano en $\mathbb{R}^3$ dado por la ecuación $x + y + z = 1$.

***EJERCICIO 1.73** Sea $\mathbf{n}$ un vector no nulo en $\mathbb{R}^n$.

a) Demuestra que $V_0 := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{n} \cdot \mathbf{x} = 0\}$ es un *subespacio vectorial* de $\mathbb{R}^n$; es decir, que la suma de vectores en V_0 se queda en V_0 y que el "alargamiento" de vectores en V_0 también está en V_0.

b) Demuestra que para cualquier $d \in \mathbb{R}$, el conjunto $V_d := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{n} \cdot \mathbf{x} = d\}$ es un trasladado de V_0; es decir, que existe un $\mathbf{p} \in \mathbb{R}^n$ tal que $V_d = V_0 + \mathbf{p} = \{\mathbf{y} + \mathbf{p} \mid \mathbf{y} \in V_0\}$.

c) ¿Qué dimensión dirías que tiene V_d? Argumenta un poco tu respuesta.

1.9 Norma y ángulos

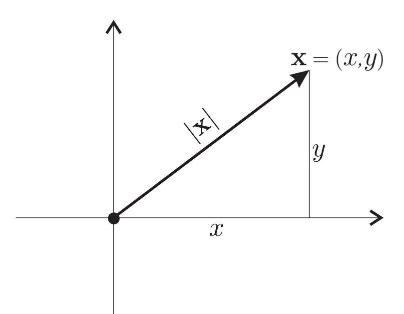
Del producto interior, obtendremos la noción de *norma* o *magnitud* de los vectores, que corresponde a la *distancia* del punto al origen.

Definición 1.9.1 Dado un vector $\mathbf{v} \in \mathbb{R}^n$, su *norma* (o *magnitud*) es el número real:

$$|\mathbf{v}| := \sqrt{\mathbf{v} \cdot \mathbf{v}},$$

de tal manera que la norma es una función de $\mathbb{R}^n$ en $\mathbb{R}$.

Como $\mathbf{v} \cdot \mathbf{v} \geq 0$ (Teorema 1.11), tiene sentido tomar su raíz cuadrada (que a su vez se define como el número positivo tal que al elevarlo al cuadrado nos da el dado). Se tiene entonces la siguiente fórmula que es una definición equivalente de la norma y que será usada con mucho más frecuencia

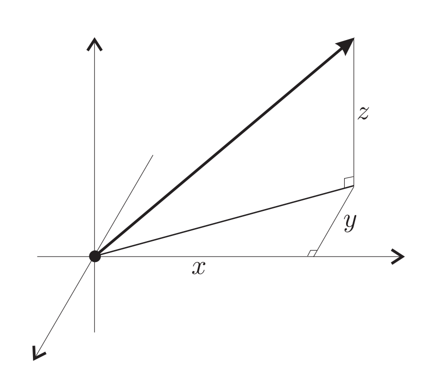


$$|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v}.$$

En $\mathbb{R}^2$ la norma se escribe, con coordenadas $\mathbf{v} = (x, y)$, como

$$|\mathbf{v}|^2 = x^2 + y^2.$$

Entonces $|\mathbf{v}|$ corresponde a la distancia euclidiana del origen al punto $\mathbf{v} = (x, y)$, pues de acuerdo con el Teorema de Pitágoras, x y y son lo que miden los catetos del triángulo rectángulo con hipotenusa $\mathbf{v}$. Aquí es donde resulta importante (por primera vez) que los ejes coordenados se tomen ortogonales, pues entonces la fórmula para calcular la distancia euclidiana al origen a partir de las coordenadas se hace sencilla.



En $\mathbb{R}^3$, con coordenadas $\mathbf{v} = (x, y, z)$, la norma se escribe

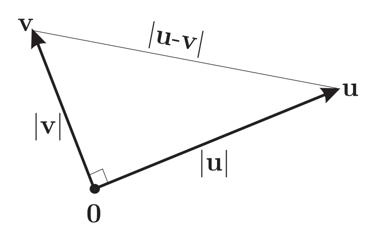
$$|\mathbf{v}|^2 = x^2 + y^2 + z^2,$$

que de nuevo, usando dos veces el Teorema de Pitágoras (en los triángulos de la figura) y que la nueva dirección z es ortogonal al plano x, y, corresponde a la *magnitud* del vector $\mathbf{v}$.

Demostremos primero el Teorema de Pitágoras para vectores en general. Este teorema fue la motivación básica para la definición de norma; sin embargo su uso ha sido sólo ése: como motivación, pues no lo hemos usado formalmente sino para ver que la norma corresponde a la noción euclidiana de distancia al origen (magnitud de vectores) cuando los ejes se toman ortogonales entre sí. Ahora veremos que al estar en el trasfondo de nuestras definiciones, éstas le hacen honor al hacerlo verdadero.

Teorema 1.20 (Pitágoras vectorial) *Sean* $\mathbf{u}$ *y* $\mathbf{v}$ *dos vectores en* $\mathbb{R}^n$*. Entonces son perpendiculares si y sólo si*

$$|\mathbf{u}|^2 + |\mathbf{v}|^2 = |\mathbf{u} - \mathbf{v}|^2 .$$



Demostración. De las propiedades del producto interior (y la definición de norma) se obtiene

$$\begin{aligned} |\mathbf{u} - \mathbf{v}|^2 &= (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) \\ &= \mathbf{u} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} - \mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} \\ &= \mathbf{u} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} - 2(\mathbf{u} \cdot \mathbf{v}) \\ &= |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2(\mathbf{u} \cdot \mathbf{v}) . \end{aligned}$$

Por definición, $\mathbf{u}$ y $\mathbf{v}$ son perpendiculares si y sólo si $\mathbf{u} \cdot \mathbf{v} = 0$ y entonces el teorema se sigue de la igualdad anterior. □

Así que nuestra definición de perpendicularidad (que el producto punto se anule) corresponde a que el Teorema de Pitágoras se vuelva cierto. Y entonces ahora tiene sentido nuestra manera informal de llamar, en $\mathbb{R}^3$, a las soluciones de la ecuación $\mathbf{u}\cdot\mathbf{x}=0$ "el plano normal a $\mathbf{u}$". Pues $\mathbf{u}\cdot\mathbf{x}=0$ si y solo si los vectores $\mathbf{u}$ y $\mathbf{x}$ son los catetos de un triángulo que cumple el Teorema de Pitágoras y por tanto es rectángulo.

Las propiedades básicas de la norma se resumen en el siguiente teorema.

Teorema 1.21 *Para todos los vectores* $\mathbf{v},\mathbf{u}\in\mathbb{R}^n$ *y para todo número real* $t\in\mathbb{R}$ *se tiene que:*

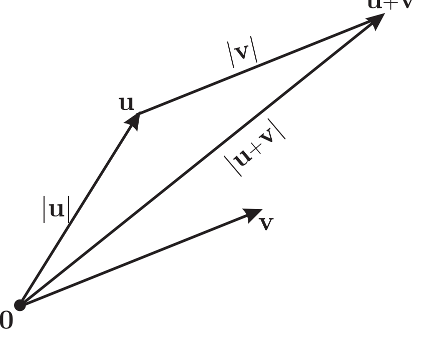
$$\begin{aligned} &\textit{i)} && |\mathbf{v}|\geq 0\\ &\textit{ii)} && |\mathbf{v}|=0\Leftrightarrow \mathbf{v}=0\\ &\textit{iii)} && |\,t\mathbf{v}\,|=|t|\,|\mathbf{v}|\\ &\textit{iv)} && |\mathbf{u}|+|\mathbf{v}|\geq|\mathbf{u}+\mathbf{v}|\\ &\textit{v)} && |\mathbf{u}|\;|\mathbf{v}|\geq|\mathbf{u}\cdot\mathbf{v}| \end{aligned}$$

Demostración. La primera afirmacion es consecuencia inmediata de la definición de raíz cuadrada, y la segunda del Teorema 1.11. La tercera, donde también se usa ese teorema, es muy simple pero con una sutileza:

$$|\,t\mathbf{v}\,|^2=(t\mathbf{v})\cdot(t\mathbf{v})=t^2(\mathbf{v}\cdot\mathbf{v})=t^2\,|\mathbf{v}|^2$$

de donde, tomando raíz cuadrada, se deduce $|\,t\mathbf{v}\,|=|t|\,|\mathbf{v}|$. Puesto que $\sqrt{t^2}$ no siempre es t sino su *valor absoluto* $|t|$, que es positivo siempre (y que podría definirse $|t|:=\sqrt{t^2}$); nótese además que como $|\mathbf{v}|\geq 0$ entonces $\sqrt{|\mathbf{v}|^2}=|\mathbf{v}|$. Obsérvese que para $n=1$ (es decir, en $\mathbb{R}$) la norma y el valor absoluto coinciden, así que usar la misma notación para ambos no causa ningún conflicto.

El inciso (**iv**) se conoce como "**la desigualdad del triángulo**", pues dice que en el triángulo con vértices 0, $\mathbf{u}$ y $\mathbf{u}+\mathbf{v}$, es más corto ir directamente de 0 a $\mathbf{u}+\mathbf{v}$ ($|\mathbf{u}+\mathbf{v}|$) que pasar primero por $\mathbf{u}$ ($|\mathbf{u}|+|\mathbf{v}|$). Para demostrarla, nótese primero que como ambos lados de la desigualdad son no negativos, entonces ésta es equivalente a que la misma desigualdad se cumpla para los cuadrados, *i.e.*,



$$|\mathbf{u}|+|\mathbf{v}|\geq|\mathbf{u}+\mathbf{v}| \quad\Leftrightarrow\quad (|\mathbf{u}|+|\mathbf{v}|)^2\geq|\mathbf{u}+\mathbf{v}|^2\,.$$

Y esta última desigualdad es equivalente a que $(|\mathbf{u}|+|\mathbf{v}|)^2-|\mathbf{u}+\mathbf{v}|^2\geq 0$. Así que debemos desarrollar el lado izquierdo:

$$\begin{aligned}(|\mathbf{u}|+|\mathbf{v}|)^2-|\mathbf{u}+\mathbf{v}|^2 &= (|\mathbf{u}|+|\mathbf{v}|)^2-(\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})\\ &= |\mathbf{u}|^2+|\mathbf{v}|^2+2\,|\mathbf{u}|\,|\mathbf{v}|-(\mathbf{u}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v}+2\,(\mathbf{u}\cdot\mathbf{v}))\\ &= 2(|\mathbf{u}|\,|\mathbf{v}|-(\mathbf{u}\cdot\mathbf{v})).\end{aligned}$$

Queda entonces por demostrar que $|\mathbf{u}|\,|\mathbf{v}| \geq (\mathbf{u}\cdot\mathbf{v})$, pero esto se sigue del inciso **(v)**, que es una afirmación más fuerte que demostraremos a continuación independientemente.

El inciso **(v)** se conoce como "**la desigualdad de Schwartz**". Por un razonamiento análogo al del inciso anterior, bastará demostrar que $(|\mathbf{u}|\,|\mathbf{v}|)^2 - |\mathbf{u}\cdot\mathbf{v}|^2$ es positivo. Lo haremos para $\mathbb{R}^2$ con coordenadas, dejando el caso de $\mathbb{R}^3$, y de $\mathbb{R}^n$, como ejercicios. Supongamos entonces que $\mathbf{u} = (a, b)$ y $\mathbf{v} = (\alpha, \beta)$ para obtener

$$\begin{aligned} |\mathbf{u}|^2\,|\mathbf{v}|^2 - |\mathbf{u}\cdot\mathbf{v}|^2 &= \left(a^2+b^2\right)\left(\alpha^2+\beta^2\right) - (a\alpha + b\beta)^2 \\ &= a^2\alpha^2 + b^2\beta^2 + a^2\beta^2 + b^2\alpha^2 \\ &\quad - \left(a^2\alpha^2 + b^2\beta^2 + 2a\alpha b\beta\right) \\ &= a^2\beta^2 - 2\,(a\beta)\,(b\alpha) + b^2\alpha^2 \\ &= (a\beta - b\alpha)^2 \geq 0\ ; \end{aligned}$$

donde la última desigualdad se debe a que el cuadrado de cualquier número es no negativo. Lo cual demuestra la desigualdad de Schwartz, la del Triángulo y completa el Teorema. □

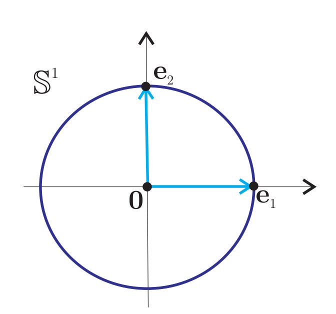
EJERCICIO 1.74 Demuestra que dos vectores $\mathbf{u}$ y $\mathbf{v}$ son perpendiculares si y sólo si $|\mathbf{u}+\mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2$. Haz el dibujo.

EJERCICIO 1.75 Demuestra que dos vectores $\mathbf{u}$ y $\mathbf{v}$ son perpendiculares si y sólo si $|\mathbf{u}+\mathbf{v}| = |\mathbf{u}-\mathbf{v}|$. Haz el dibujo. Observa que este enunciado corresponde a que un paralelogramo es un rectángulo si y sólo si sus diagonales miden lo mismo.

EJERCICIO 1.76 Sean $\mathbf{u}$ y $\mathbf{v}$ dos vectores no nulos. Demuestra que tienen la misma norma si y sólo si $\mathbf{u}+\mathbf{v}$ y $\mathbf{u}-\mathbf{v}$ son ortogonales.

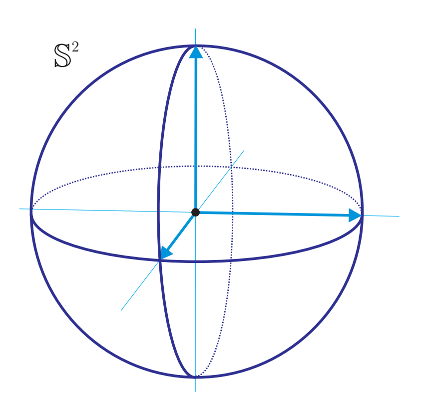
EJERCICIO 1.77 Demuestra la desigualdad de Schwartz en $\mathbb{R}^3$. ¿Puedes dar, o describir, la demostración en $\mathbb{R}^n$?

1.9.1 El círculo unitario



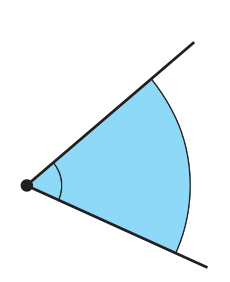
A los vectores que tienen norma igual a uno, se les llama *vectores unitarios*. Y al conjunto de todos los vectores unitarios en $\mathbb{R}^2$ se le llama el *círculo unitario* y se denota con $\mathbb{S}^1$. La notación viene de la palabra *sphere* en inglés, que significa esfera; pues en general se puede definir a la *esfera de dimensión* n como

$$\mathbb{S}^n = \left\{ \mathbf{x} \in \mathbb{R}^{n+1} \mid |\mathbf{x}| = 1 \right\}.$$

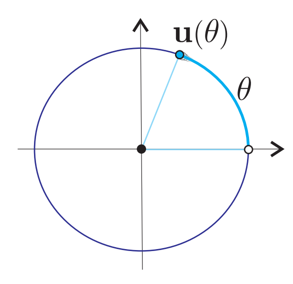


Nótese que el exponente se refiere a la dimensión de la esfera en sí, y que ésta necesita una dimensión más para "vivir". Así, la esfera de dimensión 2 vive en $\mathbb{R}^3$, y es la representación abstracta de las pompas de jabón. Más adelante la estudiaremos con cuidado. Por lo pronto nos interesa el círculo unitario.

Los puntos de $\mathbb{S}^1$ corresponden a los *ángulos*, y a estos los mediremos con radianes. La definición formal o muy precisa de ángulo involucra necesariamente nociones de cálculo que no entran en este libro. Sin embargo, es un concepto muy intuitivo y de esa intuición nos valdremos. Un *ángulo* es un sector radial del círculo unitario, aunque se denota gráficamente con un arco pequeño cerca del centro para indicar que no depende realmente del círculo de referencia, sino que es más bien un sector de cualquier círculo concéntrico. Es costumbre partir el círculo completo en 360 sectores iguales llamados *grados*. Resulta conveniente usar el número 360, pues $360 = 2^3 3^2 5$ tiene muchos divisores, así que el círculo se puede partir en cuatro sectores iguales de 90 grados (denotado $90°$ y llamado *ángulo recto*) o en 12 de $30°$ que corresponden a las horas del reloj, etc. Pero la otra manera de medirlos, que no involucra la convención de escoger 360 para la vuelta entera, es por la longitud del arco de círculo que abarcan; a esta medida se le conoce como *en radianes*. Un problema clásico que resolvieron los griegos con admirable precisión fue medir la *circunferencia* del círculo unitario (de radio uno), es decir, cuánto mide un hilo "untado" en el círculo. El resultado, como todos sabemos, es el doble del famoso número $\pi = 3.14159...$, donde los puntos suspensivos indican que su expresión decimal sigue infinitamente pues no es racional. Otra manera de entender los radianes es cinemática y es la que usaremos a continuación para establecer notación; es la versión teórica del movimiento de la piedra en una honda justo antes de lanzarla.

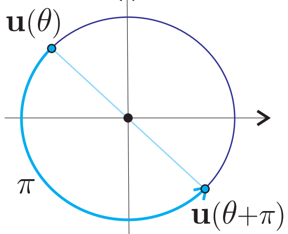


Si una partícula viaja dentro de $\mathbb{S}^1$ a velocidad constante 1, partiendo del punto $\mathbf{e}_1 = (1, 0)$ y en dirección contraria a las manecillas del reloj (es decir, saliendo hacia arriba con vector velocidad $\mathbf{e}_2 = (0, 1)$), entonces en un tiempo θ estará en un punto que llamaremos $\mathbf{u}(\theta)$; en el tiempo $\pi/2$ estará en el $\mathbf{e}_2 = (0, 1)$ (es decir, $\mathbf{u}(\pi/2) = \mathbf{e}_2$), en el tiempo π en el $(-1, 0)$ y en el 2π estará de regreso para empezar de nuevo. El tiempo aquí, que estamos midiendo con el parametro θ, corresponde precisamente a los *radianes*, pues al viajar a velocidad constante 1 el tiempo es igual a la distancia recorrida. Podemos también dar sentido a $\mathbf{u}(\theta)$ para θ negativa pensando que la partícula viene dando vueltas desde siempre (tiempo infinito negativo) de tal manera que en el tiempo 0 pasa justo por $\mathbf{e}_1$, es decir, que



$$\mathbf{u}(0) = (1, 0).$$

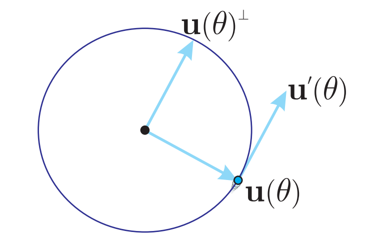
Nuestra suposición básica es que la función $\mathbf{u} : \mathbb{R} \to \mathbb{S}^1$, que hemos descrito, está bien definida. Se cumple entonces que



$$\mathbf{u}(\theta) = \mathbf{u}(\theta + 2\pi),$$

pues en el tiempo 2π la partícula da justo una vuelta (la longitud del círculo mide lo mismo independientemente de dónde empecemos a medir). Y también que $\mathbf{u}(\theta + \pi) = -\mathbf{u}(\theta)$, pues π es justo el ángulo que da media vuelta.

Observemos ahora que al pedir que la partícula viaje en el círculo unitario con velocidad constante 1 entonces su *vector velocidad*, que incluye ahora dirección además de magnitud, es tangente a $\mathbb{S}^1$ (la piedra de la honda sale por la tangente), y es fácil ver que entonces es perpendicular al vector de posición $\mathbf{u}(\theta)$, pero más precisamente es justo su compadre ortogonal porque va girando en "su" dirección, "hacia él". Usando la notación de derivadas, podríamos escribir

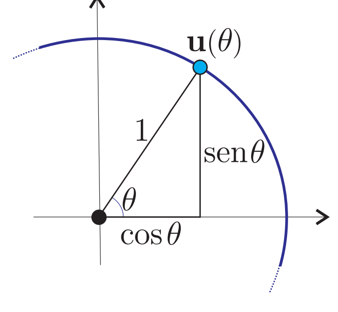


$$\mathbf{u}'(\theta) = \mathbf{u}(\theta)^{\perp}.$$

Hay que remarcar que al medir ángulos con radianes, si bien ganamos en naturalidad matemática, es inevitable la ambigüedad de que a los ángulos no corresponda un número único. Pues θ y $\theta + 2n\pi$, para cualquier $n \in \mathbb{Z}$, determinan al mismo ángulo, es decir $\mathbf{u}(\theta) = \mathbf{u}(\theta + 2n\pi)$. Podemos exigir que θ esté en el intervalo entre 0 y 2π, o bien, que resulta más agradable geométricamente, en el intervalo de $-\pi$ a π, para reducir la ambigüedad sólo a los extremos. Pero al sumar o restar ángulos, inevitablemente nos saldremos de este intervalo y habrá que volver a ajustar.

Funciones trigonométricas

Recordemos ahora que $\mathbb{S}^1$ es un subconjunto de $\mathbb{R}^2$, así que el punto $\mathbf{u}(\theta)$ tiene dos coordenadas precisas. Éstas están dadas por las importantísimas *funciones trigonométricas* coseno y seno. Más precisamente, podemos definir



$$(\cos\theta, \operatorname{sen}\theta) := \mathbf{u}(\theta),$$

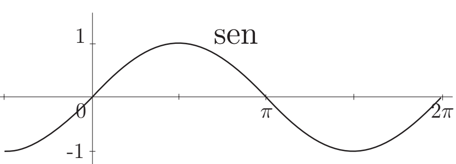
es decir, $\cos\theta$ y $\operatorname{sen}\theta$ son las coordenadas cartesianas del punto $\mathbf{u}(\theta) \in \mathbb{S}^1$. Entonces el coseno y el seno corresponden respectivamente al cateto adyacente y al cateto opuesto de un triángulo rectángulo con hipotenusa 1 y ángulo θ, como se ve en trigonometría de secundaria. También pensarse que si damos por establecidas las funciones seno y coseno, entonces la función $\mathbf{u}(\theta)$ está dada por la ecuación anterior. Hay que remarcar que la pertenencia $\mathbf{u}(\theta) \in \mathbb{S}^1$ equivale entonces a la ecuación

$$\operatorname{sen}^2\theta + \cos^2\theta = 1.$$

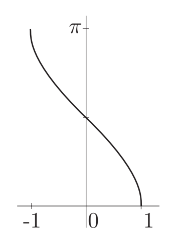
Las propiedades de la función $\mathbf{u}$ que hemos enumerado, traducidas a sus coordenadas (por ejemplo, la del vector velocidad se traduce a $\cos'\theta = -\operatorname{sen}\theta$ y $\operatorname{sen}'\theta = \cos\theta$) son suficientes para definir a las funciones cos y sen, y en adelante las daremos por un hecho. Sus bien conocidas gráficas aparecen en la figura.

Observemos, por último, que cualquiera de las funciones coseno o seno casi determinan al ángulo, pues de la ecuación anterior se pueden despejar con la ambigüedad de un signo al tomar raíz cuadrada, es decir,

$$\operatorname{sen}\theta = \pm\sqrt{1-\cos^2\theta}.$$



Dicho de otra manera, dado x en el intervalo de -1 a 1 (escrito $x \in [-1,1]$,[4] o bien $-1 \leq x \leq 1$) tenemos dos posibles puntos en $\mathbb{S}^1$ con esa ordenada, a saber $\left(x, \sqrt{1-x^2}\right)$ y $\left(x, -\sqrt{1-x^2}\right)$. Entonces con la ambigüedad de un signo podemos determinar el ángulo del cuál x es el coseno. Si escogemos la parte superior del círculo unitario obtenemos una función, llamada *arcocoseno* y leída *"el arco cuyo coseno es ..."*



$$\arccos : [-1,1] \to [0,\pi]\,,$$

y definida por

$$\cos(\arccos(x)) = x.$$

La ambigüedad surge al componer en el otro sentido, pues para $\theta \in [-\pi,\pi]$ se tiene $\arccos(\cos(\theta)) = \pm\theta$. Obsérvese que entonces la mitad superior del círculo unitario se describe como la gráfica de la función $x \mapsto \operatorname{sen}(\arccos(x))$.

EJERCICIO 1.78 Demuestra (usando que definimos $\mathbf{u}(\theta) \in \mathbb{S}^1$) que para cualquier $\theta \in \mathbb{R}$ se cumplen

$$\begin{aligned} -1 &\leq \cos\theta \leq 1 \\ -1 &\leq \operatorname{sen}\theta \leq 1 \end{aligned}$$

EJERCICIO 1.79 Demuestra que para cualquier $\theta \in \mathbb{R}$ se cumplen

$$\begin{array}{ccc} \cos(\theta+\pi/2) = -\operatorname{sen}\theta & \cos(\theta+\pi) = -\cos\theta & \cos(-\theta) = \cos\theta \\ \operatorname{sen}(\theta+\pi/2) = \cos\theta & \operatorname{sen}(\theta+\pi) = -\operatorname{sen}\theta & \operatorname{sen}(-\theta) = -\operatorname{sen}\theta \end{array}$$

EJERCICIO 1.80 Construye una tabla con los valores de las funciones cos y sen para los ángulos $0, \pi/6, \pi/4, \pi/3, \pi/2, 2\pi/3, 3\pi/4, 5\pi/6, \pi, 5\pi/4, 3\pi/2, -\pi/3, -\pi/4$, y dibuja los correspondientes vectores unitarios.

[4]Dados $a, b \in \mathbb{R}$, con $a \leq b$, denotamos por $[a,b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}$ al *intervalo cerrado* entre a y b.

EJERCICIO 1.81 ¿Cuál es el conjunto de números que usan las coordenadas de las horas del reloj?

1.9.2 Coordenadas polares

El círculo unitario $\mathbb{S}^1$ (cuyos puntos hemos identificado con los ángulos tomando su ángulo con el vector base $\mathbf{e}_1 = (1,0)$) tiene justo un representante de cada posible dirección en $\mathbb{R}^2$, donde ahora una dirección y su opuesta son diferentes (aunque sean, según nuestra definición, paralelas). De tal manera que a cualquier punto de $\mathbb{R}^2$ que no sea el origen se llega viajando en una (y exactamente una) de estas direcciones. Ésta es la idea central de las coordenadas polares, que en muchas situaciones son más naturales, o útiles, para identificar los puntos del plano. Por ejemplo, son las que intuitivamente usa un cazador: apuntar —decidir una dirección— es la primera coordenada y luego, dependiendo de la distancia a la que vuela el pato —la segunda coordenada— tiene que ajustar el tiro para que las trayectorias de pato y perdigones se intersecten.

Sea $\mathbf{x}$ cualquier vector no nulo en $\mathbb{R}^2$. Entonces $|\mathbf{x}| \neq 0$ y tiene sentido tomar el vector $\left(|\mathbf{x}|^{-1}\right)\mathbf{x}$, que es unitario pues como $|\mathbf{x}| > 0$ implica que $|\mathbf{x}|^{-1} > 0$, entonces tenemos

$$\left||\mathbf{x}|^{-1}\mathbf{x}\right| = \left||\mathbf{x}|^{-1}\right||\mathbf{x}| = |\mathbf{x}|^{-1}|\mathbf{x}| = 1.$$

De aquí que exista un único $\mathbf{u}(\theta) \in \mathbb{S}^1$ (aunque θ no es único como real, sí lo es como ángulo), tal que

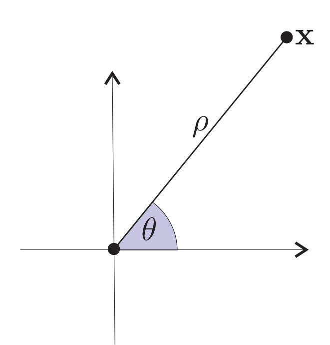
$$\mathbf{u}(\theta) = |\mathbf{x}|^{-1}\mathbf{x}$$

y claramente se cumple que

$$\mathbf{x} = |\mathbf{x}|\,\mathbf{u}(\theta).$$

A $\mathbf{u}(\theta)$ se le referirá como la *dirección* de $\mathbf{x}$. A la pareja $(\theta, |\mathbf{x}|)$ se le llama las *coordenadas polares* del vector $\mathbf{x}$. Se usa el término "coordenadas" pues determinan al vector, es decir, si nos dan una pareja (θ, ρ) con $\rho \geq 0$, obtenemos un vector

$$\mathbf{x} = \rho\,\mathbf{u}(\theta)$$



con magnitud ρ y dirección $\mathbf{u}(\theta)$, o bien, *ángulo* θ. Obsérvese que sólo con el origen hay ambigüedad: si $\rho = 0$ en la formúla anterior, para cualquier valor de θ obtenemos el origen; éste no tiene un ángulo definido. Así que al hablar de coordenadas polares supondremos que el vector en cuestión es no nulo. Por nuestra convención de $\mathbf{u}(\theta)$, lo que representa θ es el ángulo respecto al eje x en su dirección positiva.

EJERCICIO 1.82 Da las coordenadas (cartesianas) de los vectores cuyas coordenadas polares son $(\pi/2, 3)$, $\left(\pi/4, \sqrt{2}\right)$, $(\pi/6, 2)$, $(\pi/6, 1)$.

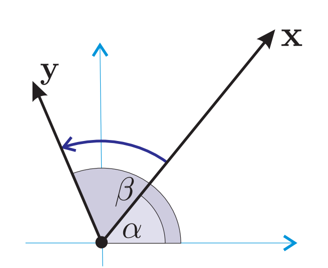
EJERCICIO 1.83 Si tomamos (θ, ρ) como coordenadas polares, describe los subconjuntos de $\mathbb{R}^2$ definidos por las ecuaciones $\theta = \mathtt{cte}$ ("θ igual a una constante") y $\rho = \mathtt{cte}$.

1.9.3 Ángulo entre vectores

Podemos usar coordenadas polares para definir el ángulo entre vectores en general.

Definición 1.9.2 Dados x y y, dos vectores no nulos en $\mathbb{R}^2$, sean $(\alpha, |x|)$ y $(\beta, |y|)$ sus coordenadas polares respectivamente. El *ángulo de* x *a* y es

$$\overrightarrow{\text{ang}}(x, y) = \beta - \alpha.$$



Obsérvese que el ángulo tiene dirección (de ahí que le hayamos puesto la flechita de sombrero), ya que claramente

$$\overrightarrow{\text{ang}}(x, y) = -\overrightarrow{\text{ang}}(y, x).$$

Pues $\overrightarrow{\text{ang}}(x, y)$ corresponde al movimiento angular que nos lleva de la dirección de x a la de y, y $\overrightarrow{\text{ang}}(y, x)$ es justo el movimiento inverso. Así que podemos resumir diciendo que las coordenadas polares de un vector x (recúerdese que al aplicar el término ya suponemos que es no nulo) es la pareja

$$\left(\overrightarrow{\text{ang}}(e_1, x), |x|\right).$$

Con mucha frecuencia van a aparecer ángulos no dirigidos, pensados como un sector del círculo unitario sin principio ni fin determinados. Para ellos usaremos la notación $\text{ang}(x, y)$ (sin la flecha), llamándolo el *ángulo* <u>entre</u> x y y. Aunque intuitivamente es claro lo que queremos (piénsese en una porción de una pizza), su definición es un poco más laboriosa. Si tomamos dos vectores unitarios en $\mathbb{S}^1$, estos parten el círculo en dos sectores, uno de ellos es menor que medio círculo y ése es el ángulo entre ellos. Así que debemos pedir que

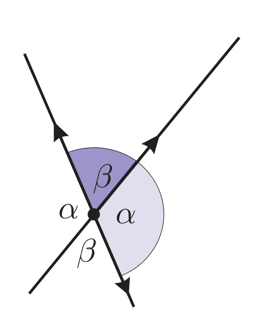
$$0 \leq \text{ang}(x, y) \leq \pi.$$

Y es claro que escogiendo a α y a β (los ángulos de x y y, respectivamente) de manera adecuada se tiene que $\text{ang}(x, y) := |\beta - \alpha| \in [0, \pi]$. Así que, por ejemplo, $\text{ang}(x, y) = 0$ significa que x y y apuntan en la misma dirección, mientras que $\text{ang}(x, y) = \pi$ significa que x y y apuntan en direcciones contrarias, y se tiene que $\text{ang}(x, y) = \pi/2$ cuando son perpendiculares.

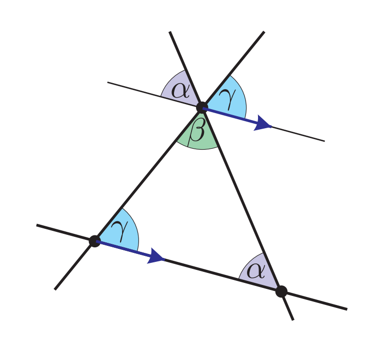
EJERCICIO 1.84 ¿Cuál es el ángulo de x a y, cuando a) $x = (1, 0)$, $y = (0, -2)$; b) $x = (1, 1)$, $y = (0, 1)$; c) $x = \left(\sqrt{3}, 2\right)$, $y = (1, 0)$.

Suma de ángulos

Si ya definimos ángulo entre vectores, resulta natural definir *ángulo entre dos rectas* como el ángulo entre sus vectores direccionales. Pero hay una ambigüedad pues podemos escoger direcciones opuestas para una misma recta. Esta ambigüedad equivale a que dos rectas que se intersectan definen cuatro sectores o "ángulos" que se agrupan naturalmente en dos parejas opuestas por el vértice que miden lo mismo; y estos dos ángulos son *complementarios*, es decir, suman π. No vale la pena entrar en los detalles formales de esto usando vectores direccionales pues sólo se confunde lo obvio, que desde muy temprana edad sabemos. Pero sí podemos delinear rápidamente la demostración de que la suma de los ángulos de un triángulo siempre es π, que para Euclides tuvo que ser axioma (disfrazado o no, pues recuérdese que es equivalente al Quinto).



Un triángulo son tres rectas que se intersectan dos a dos (pero no las tres), y en cada vértice (la intersección de dos de las rectas) define un ángulo interno; llamémoslos α, β y γ. Al trazar la paralela a un lado del triángulo que pasa por el vértice opuesto (usando el Quinto), podemos medir ahí los tres ángulos, pues por nuestra demostración del Quinto esta paralela tiene el mismo vector direccional que el lado original, y nuestra definición de ángulo entre líneas usa la de ángulo entre vectores direccionales. Finalmente, hay que observar que si en ese vértice cambiamos uno de los ángulos por su opuesto, los tres ángulos se juntan para formar el ángulo de un vector a su opuesto, es decir, suman π. Hemos delineado la demostración del siguiente teorema clásico.

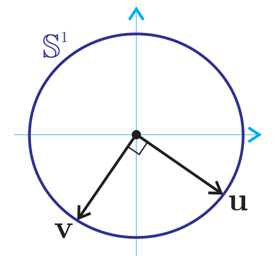


Teorema 1.22 *Los ángulos internos de un triángulo suman* π. □

1.10 Bases ortonormales

Definición 1.10.1 Una *base ortonormal* de $\mathbb{R}^2$ es una pareja de vectores unitarios perpendiculares.

Por ejemplo, la *base canónica* $\mathbf{e}_1 = (1, 0)$, $\mathbf{e}_2 = (0, 1)$ es una base ortonormal, y para cualquier $\mathbf{u} \in \mathbb{S}^1$ se tiene que $\mathbf{u}, \mathbf{u}^\perp$ es una base ortonormal, pues también $|\mathbf{u}^\perp| = 1$. Obsérvese además que si $\mathbf{u}$ y $\mathbf{v}$ son una base ortonormal entonces ambos están en $\mathbb{S}^1$ y además $\mathbf{v} = \mathbf{u}^\perp$ o $\mathbf{v} = -\mathbf{u}^\perp$. Por el momento, su importancia radica en que es muy fácil escribir cualquier vector como combinación lineal de una base ortonormal:

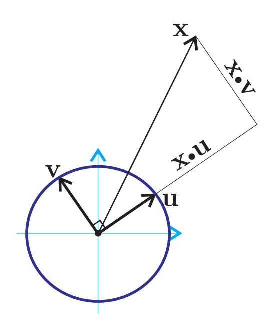


Teorema 1.23 *Sean* $\mathbf{u}$ *y* $\mathbf{v}$ *una base ortonormal de* $\mathbb{R}^2$*, entonces para cualquier* $\mathbf{x} \in \mathbb{R}^2$ *se tiene que*

$$\mathbf{x} = (\mathbf{x} \cdot \mathbf{u})\,\mathbf{u} + (\mathbf{x} \cdot \mathbf{v})\,\mathbf{v}.$$

Demostración. Suponemos que $\mathbf{x}$ está dado, y vamos a resolver el sistema de ecuaciones

$$s\,\mathbf{u} + t\,\mathbf{v} = \mathbf{x} \tag{1.11}$$



con incógnitas t, s. Tomando el producto interior con $\mathbf{u}$ en la ecuación anterior obtenemos

$$s\,(\mathbf{u} \cdot \mathbf{u}) + t\,(\mathbf{v} \cdot \mathbf{u}) = \mathbf{x} \cdot \mathbf{u}.$$

Pero $\mathbf{u} \cdot \mathbf{u} = 1$ pues $\mathbf{u}$ es unitario y $\mathbf{v} \cdot \mathbf{u} = 0$ ya que $\mathbf{u}$ y $\mathbf{v}$ son perpendiculares, así que

$$s = \mathbf{x} \cdot \mathbf{u}.$$

Análogamente, tomando el producto interior por $\mathbf{v}$, se obtiene que $t = \mathbf{x} \cdot \mathbf{v}$. De tal manera que el sistema (1.11) sí tiene solución y es la que asegura el teorema. □

Vale la pena observar que el truco que acabamos de usar para resolver un sistema de dos ecuaciones con dos incógnitas no es nuevo, lo usamos en la Sección 1.7.1 al tomar el producto interior con el compadre ortogonal; la diferencia es que aquí el otro vector de la base ortonormal está dado.

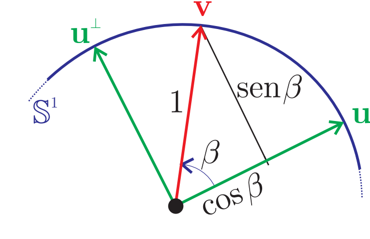
Como corolario de este teorema obtenemos la interpretación geométrica del producto interior de vectores unitarios.

Corolario 1.24 *Si* $\mathbf{u}$, $\mathbf{v} \in \mathbb{S}^1$ *y* α *es el ángulo entre ellos,* i.e., $\alpha = \operatorname{ang}(\mathbf{u}, \mathbf{v})$, *entonces*

$$\mathbf{u} \cdot \mathbf{v} = \cos\alpha.$$

Demostración. Sea $\beta = \overrightarrow{\operatorname{ang}}(\mathbf{u}, \mathbf{v})$ el ángulo de $\mathbf{u}$ a $\mathbf{v}$, de tal manera que, escogiéndolos adecuadamente, se tiene que $\alpha = \pm\beta$. Usando el teorema anterior con la base ortonormal $\mathbf{u}, \mathbf{u}^\perp$ y el vector $\mathbf{v}$ (en vez de $\mathbf{x}$) se obtiene

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{u})\,\mathbf{u} + (\mathbf{v} \cdot \mathbf{u}^\perp)\,\mathbf{u}^\perp.$$



Pero también es claro geométricamente que

$$\mathbf{v} = \cos\beta\,\mathbf{u} + \operatorname{sen}\beta\,\mathbf{u}^\perp.$$

Puesto que $\cos\beta = \cos\alpha$, pues $\alpha = \pm\beta$, el corolario se sigue de que la solución de un sistema con determinante no cero es única (en nuestro caso el determinante del sistema $\mathbf{v} = s\,\mathbf{u} + t\,\mathbf{u}^\perp$ es $\det\left(\mathbf{u}, \mathbf{u}^\perp\right) = \mathbf{u}^\perp \cdot \mathbf{u}^\perp = 1$). □

EJERCICIO 1.85 Demuestra que si $\mathbf{u}$, $\mathbf{v}$ y $\mathbf{w}$ son una *base ortonormal* de $\mathbb{R}^3$ (es decir, si los tres son unitarios y dos a dos son perpendiculares), entonces para cualquier $\mathbf{x} \in \mathbb{R}^3$ se tiene que

$$\mathbf{x} = (\mathbf{x} \cdot \mathbf{u})\,\mathbf{u} + (\mathbf{x} \cdot \mathbf{v})\,\mathbf{v} + (\mathbf{x} \cdot \mathbf{w})\,\mathbf{w}.$$

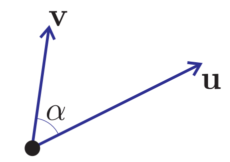
EJERCICIO 1.86 Escribe los vectores $(1, 0)$, $(0, 1)$, $(2, 1)$ y $(-1, 3)$ como combinación lineal de $\mathbf{u} = (3/5, 4/5)$ y $\mathbf{v} = (4/5, -3/5)$, es decir, como $t\,\mathbf{u} + s\,\mathbf{v}$ con t y s apropiadas.

1.10.1 Fórmula geométrica del producto interior

Como consecuencia del corolario anterior obtenemos un resultado importante que nos dice que el producto interior puede definirse en términos de la norma y el ángulo entre vectores.

Teorema 1.25 *Sean* $\mathbf{u}$ *y* $\mathbf{v}$ *dos vectores en* $\mathbb{R}^2$ *y* α *el ángulo entre ellos, entonces*

$$\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}|\,|\mathbf{v}| \cos \alpha.$$



Demostración. Si $\mathbf{u} = 0$ (o $\mathbf{v} = 0$) la igualdad anterior se cumple (pues ambos lados son 0); podemos suponer entonces que ambos son distintos de cero. Ahora podemos reescalar a $\mathbf{u}$ (y a $\mathbf{v}$) para ser unitarios (pues tiene sentido $(|\mathbf{u}|^{-1})\mathbf{u}$ que es un vector unitario) manteniendo el ángulo entre ellos. Y por el Corolario 1.24,

$$\cos \alpha = (\frac{1}{|\mathbf{u}|})\mathbf{u} \cdot (\frac{1}{|\mathbf{v}|})\mathbf{v} = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}|\ |\mathbf{v}|},$$

de donde se sigue imediatamente el teorema. □

De esta fórmula geométrica para el producto interior, se obtiene que $|\mathbf{u} \cdot \mathbf{v}| = |\cos \alpha|\,|\mathbf{u}|\ |\mathbf{v}|$, así que la desigualdad de Schwartz ($|\mathbf{u} \cdot \mathbf{v}| \geq |\mathbf{u}|\ |\mathbf{v}|$) es equivalente a $|\cos \alpha| \leq 1$.

EJERCICIO 1.87 Demuestra *la ley de los cosenos* para vectores, es decir, que dados dos vectores $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ con ángulo α entre ellos, se cumple que

$$|\mathbf{u} - \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2\,|\mathbf{u}|\,|\mathbf{v}| \cos \alpha.$$

Observa que el Teorema de Pitágoras es un caso especial.

EJERCICIO 1.88 Encuentra la distancia de los puntos $\mathbf{p}_1 = (0, 5)$, $\mathbf{p}_2 = (4, 0)$ y $\mathbf{p}_3 = (-1, 1)$ a la recta $\ell = \{t(4, 3) \mid t \in \mathbb{R}\}$.

1.10.2 El caso general

Puesto que el Teorema 1.23 y su Corolario 1.24 se demostraron para $\mathbb{R}^2$, puede dudarse de la validez de los resultados en el caso general de $\mathbb{R}^n$, pero nos interesa establecer el caso de $n = 3$. Vale la pena entonces hacer unas aclaraciones y ajustes a manera de repaso. Lo que realmente se uso del Teorema 1.23 fue que cuando $\mathbf{u}$ es un vector unitario (en cualquier dimensión ahora) entonces $\mathbf{x} \cdot \mathbf{u}$ tiene un significado geométrico muy preciso: es lo que hay que viajar en la dirección $\mathbf{u}$ para que de ahí, $\mathbf{x}$ quede en dirección ortogonal a $\mathbf{u}$. Podemos precisarlo de la siguiente manera.

Lema 1.26 *Sea $\mathbf{u} \in \mathbb{R}^n$ un vector unitario (es decir, tal que $|\mathbf{u}| = 1$). Entonces para cualquier $\mathbf{x} \in \mathbb{R}^n$ existe un vector $\mathbf{y}$ perpendicular a $\mathbf{u}$ y tal que*

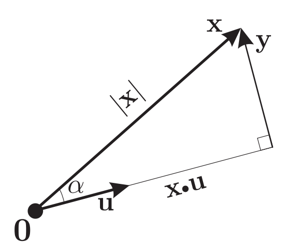
$$\mathbf{x} = (\mathbf{x} \cdot \mathbf{u})\, \mathbf{u} + \mathbf{y}.$$

Demostración. Es muy fácil, declaremos $\mathbf{y} := \mathbf{x} - (\mathbf{x} \cdot \mathbf{u})\, \mathbf{u}$ y basta ver que $\mathbf{y} \cdot \mathbf{u} = 0$:

$$\begin{aligned} \mathbf{y} \cdot \mathbf{u} &= (\mathbf{x} - (\mathbf{x} \cdot \mathbf{u})\, \mathbf{u}) \cdot \mathbf{u} \\ &= \mathbf{x} \cdot \mathbf{u} - (\mathbf{x} \cdot \mathbf{u})(\mathbf{u} \cdot \mathbf{u}) = \mathbf{x} \cdot \mathbf{u} - \mathbf{x} \cdot \mathbf{u} = 0\,. \end{aligned}$$

□

Si pensamos ahora en el plano generado por $\mathbf{u}$ y $\mathbf{x}$ (en el que también se encuentra $\mathbf{y}$) y dentro de él en el triángulo rectángulo $\mathbf{0}, (\mathbf{x} \cdot \mathbf{u})\, \mathbf{u}, \mathbf{x}$, se obtiene que



$$\cos\alpha = \frac{\mathbf{x} \cdot \mathbf{u}}{|\mathbf{x}|},$$

donde α es el ángulo entre los vectores $\mathbf{u}$ y $\mathbf{x}$. Y de aquí, el Teorema 1.25 para $\mathbb{R}^n$ se sigue inmediatamente al permitir que $\mathbf{u}$ sea no necesariamente unitario.

Si nos ponemos quisquillosos con la formalidad, se notará una falla en lo anterior (pasamos un *strike* al lector), pues no hemos definido "ángulo entre vectores" más allá de $\mathbb{R}^2$. Tómese entonces como la motivación intuitiva para la definición general que hará que todo cuadre bien.

Definición 1.10.2 Dados dos vectores no nulos $\mathbf{u}$ y $\mathbf{v}$ en $\mathbb{R}^n$ el *ángulo entre ellos* es

$$\text{ang}\,(\mathbf{u}, \mathbf{v}) = \arccos\left(\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}|\,|\mathbf{v}|}\right).$$

Nótese que por la desigualdad de Schwartz (dejada al lector para $\mathbb{R}^3$ en el Ejercicio 1.77) el argumento de la función arccos está en el intervalo $[-1, 1]$, así que está bien definido el ángulo, y queda en el intervalo $[0, \pi]$.

Hay que remarcar que a partir de $\mathbb{R}^3$ ya no tiene sentido hablar de ángulos dirigidos; esto sólo se puede hacer en el plano. Pues si tomamos dos vectores no nulos

$\mathbf{u}$ y $\mathbf{v}$ en $\mathbb{R}^3$, no hay manera de decir si el viaje angular de $\mathbf{u}$ hacia $\mathbf{v}$ es positivo o negativo. Esto dependería de escoger un lado del plano para "verlo" y ver si va contra o con el reloj, pero cualquiera de los dos lados son "iguales" y desde cada uno se "ve" lo contrario del otro; no hay una manera coherente de decidir. Lo que sí se puede decir es cuándo *tres* vectores están orientados positivamente, pero esto se verá mucho más adelante. De tal manera que a partir de $\mathbb{R}^3$, sólo el ángulo *entre* vectores está definido, y tiene valores entre 0 y π; donde los extremos corresponden a paralelismo pero con la misma dirección (ángulo 0) o la contraria (ángulo π).

EJERCICIO 1.89 Sean $\mathbf{u}$ y $\mathbf{v}$ dos vectores unitarios y ortogonales en $\mathbb{R}^3$. Demuestra que para cualquier $\mathbf{x} \in \mathbb{R}^3$ existe un único $\mathbf{w} \in \langle \mathbf{u}, \mathbf{v} \rangle$ (el plano generado por $\mathbf{u}$ y $\mathbf{v}$) tal que $\mathbf{x} - \mathbf{w}$ es ortogonal a $\mathbf{u}$ y a $\mathbf{v}$; al punto $\mathbf{w}$ se le llama la *proyección ortogonal* de $\mathbf{x}$ al plano $\langle \mathbf{u}, \mathbf{v} \rangle$. ¿Puedes concluir que la distancia del punto $\mathbf{x}$ al plano $\langle \mathbf{u}, \mathbf{v} \rangle$ es

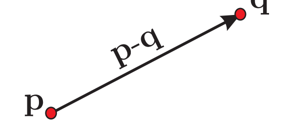
$$\sqrt{|\mathbf{x}|^2 - (\mathbf{x} \cdot \mathbf{u})^2 - (\mathbf{x} \cdot \mathbf{v})^2} \ ?$$

1.11 Distancia

Esta sección cierra el capítulo con una breve discusión sobre el concepto euclidiano de distancia, que se deduce naturalmente de la norma

Dados dos puntos $\mathbf{p}$, $\mathbf{q} \in \mathbb{R}^n$ se puede definir su *distancia euclidiana*, o simplemente su *distancia*, $d(\mathbf{p}, \mathbf{q})$, como la norma de su diferencia, es decir, como la magnitud del vector que lleva a uno en otro, *i.e.*,

$$d(\mathbf{p}, \mathbf{q}) = |\mathbf{p} - \mathbf{q}|,$$



que explícitamente en coordenadas da la fórmula (en $\mathbb{R}^2$)

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

Las propiedades básicas de la distancia se reúnen en el siguiente:

Teorema 1.27 *Para todos los vectores* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ *se cumple que*

i) $d(\mathbf{x}, \mathbf{y}) \geq 0$
ii) $d(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$
iii) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$
iv) $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$.

El espacio métrico es un conjunto en el que está definida una distancia que cumple con las cuatro propiedades del Teorema 1.27.

EJERCICIO 1.90 Da explícitamente con coordenadas la fórmula de la distancia en $\mathbb{R}^3$ y en $\mathbb{R}^n$.

EJERCICIO 1.91 Demuestra el Teorema 1.27.

EJERCICIO 1.92 ¿Podrías mencionar algún espacio métrico distinto del que se definió arriba?

1.11.1 El espacio euclidiano (primera misión cumplida)

Cuando $\mathbb{R}^n$ se considera junto con la distancia antes definida (llamada la *distancia euclidiana*), se dice que es el espacio euclidiano de dimensón n —que a veces se denota $\mathbb{E}^n$, y que formalmente podríamos definir $\mathbb{E}^n := (\mathbb{R}^n, d)$— ya que éste cumple con los postulados de Euclides para $n = 2$. Veamos.

Hemos demostrado que $\mathbb{R}^2$ cumple los tres postulados que se refieren a las rectas, y aún nos falta analizar dos. El II se refiere a "trazar" círculos. Ya hemos trabajado con el círculo unitario y claramente se generaliza. Dado un punto $\mathbf{p}$ y una distancia r (un número positivo) definimos el *círculo con centro* $\mathbf{p}$ *y radio* r como el conjunto de puntos a distancia r de $\mathbf{p}$ (que se estudiará en el siguiente capítulo). Se trata de cierto subconjunto de $\mathbb{R}^2$, que además es no vacío (es fácil dar explícitamente un punto en él usando coordenadas); y nuestra noción de "definir" un conjunto es la análoga de "trazar" para los griegos. Se cumple entonces el axioma II.

Nos falta únicamente discutir el IV, "todos los ángulos rectos son iguales". Esta afirmación es sutil. Tenemos, desde hace un buen rato, la noción de ángulo recto, perpendicularidad u ortogonalidad, pero ¿a qué se refiere la palabra "igualdad" en el contexto griego? Si fuera a la igualdad del numerito que mide los ángulos, la afirmación es obvia, casi vacua —"$\pi/2$ es igual a $\pi/2$"—, y no habría necesidad de enunciarla como axioma. Entonces se refiere a algo mucho más profundo. En este postulado está implícita la noción de *movimiento*; se entiende la "igualdad" como que podemos *mover* el plano para llevar un ángulo recto formado por dos rectas a cualquier otro. Esto es claro al ver otros teoremas clásicos como "dos triángulos son *iguales* si sus lados miden lo mismo": no se puede referir a la igualdad estricta de conjuntos, quiere decir que se puede *llevar* a uno sobre el otro para que, entonces sí, coincidan como conjuntos. Y este *llevar* es el *movimiento* implícito en el uso de la palabra "igualdad". En términos modernos, esta noción de *movimiento* del plano se formaliza con la noción de función. Nosotros lo haremos en el capítulo 3, y corresponderá formalmente a la noción de *isometría* (una función de $\mathbb{R}^2$ en $\mathbb{R}^2$ que preserva distancias). Pero ya no queremos distraernos con la discusión de la axiomática griega. Creamos de buena fe que al desarrollar formalmente los conceptos necesarios el Postulado IV será cierto, o tomémoslo en su acepción numérica trivial, para concluir esta discusión con la que arrancamos el libro.

A partir de los axiomas de los números reales construimos el Plano Euclidiano,

$\mathbb{E}^2$, con todo y sus nociones básicas que cumplen los postulados de Euclides. Se tiene entonces que para estudiarlo son igualmente válidos el método *sintético* (el que usaban los griegos pues sus axiomas ya son ciertos) o el *analítico* (que inauguró Descartes y estamos siguiendo). En este último método siempre hay elementos del primero, no siempre es más directo irse a las coordenadas; no hay un divorcio y en general es muy difícil trazar la línea que los separa; además no vale la pena. No hay que preocuparse de eso, en cada momento toma uno lo que le conviene. Pero sí hay que hacer énfasis en la enorme ventaja que dio el método cartesiano, al construir de golpe (y sólo con un poco de esfuerzo extra) espacios euclidianos para todas las dimensiones. Es un método que permitió generalizar y abrir, por tanto, nuevos horizontes. Y no sólo en cuestión de dimensiones.

Como se verá en capítulos posteriores, siguiendo el método analítico pueden construirse espacios de dimensión 2 con formas "raras" de medir ángulos y distancias que los hacen cumplir todos los axiomas menos el Quinto, y haciéndolos entonces espacios *no euclidianos*. Pero esto vendrá a su tiempo; por el momento y para cerrar el círculo de esta discusión, reescribiremos con la noción de distancia el Teorema 1.1, cuya demostración se deja como ejercicio, y donde, nótese, tenemos el extra del "si y sólo si".

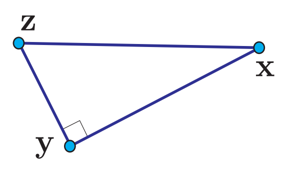
Se puede definir ángulo en tercias ordenadas de puntos. De nuevo, refiriéndonos al ángulo que ya definimos entre vectores (haciendo que el punto de enmedio sea como el origen):

$$\angle xyz := \text{ang}\left((x-y),(z-y)\right) = \arccos\frac{(x-y)\cdot(z-y)}{|x-y|\,|z-y|}.$$

Y entonces tenemos:

Teorema 1.28 (Pitágoras) *Dados tres puntos x, y, z en $\mathbb{E}^n$, se tiene que*

$$d(x,\, z)^2 = d(x,\, y)^2 + d(y,\, z)^2 \quad \Leftrightarrow \quad \angle xyz = \pi/2.$$



□

Para retomar el hilo de la corriente principal del texto, en los siguientes ejercicios vemos cómo con la noción de distancia se pueden reconstruir objetos básicos como segmentos y rayos, que juntos dan las líneas; y delineamos un ejemplo de un espacio métrico que da un plano con una geometría "rara".

EJERCICIO 1.93 Demuestra esta versión del Teorema de Pitágoras.

EJERCICIO 1.94 Demuestra que el segmento de x a y es el conjunto

$$\overline{xy} = \{z \mid d(x,y) = d(x,z) + d(z,y)\}.$$

EJERCICIO 1.95 Demuestra que el *rayo que parte de* $\mathbf{y}$ *desde* $\mathbf{x}$ (es decir la continuación de la recta por $\mathbf{x}$ y $\mathbf{y}$ más allá de $\mathbf{y}$) es el conjunto $\{\mathbf{z} \mid d(\mathbf{x},\mathbf{y}) + d(\mathbf{y},\mathbf{z}) = d(\mathbf{x},\mathbf{z})\}$.

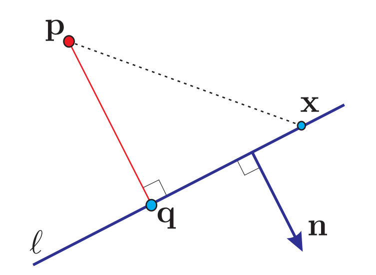
***EJERCICIO 1.96** (Un ejemplo cotidiano de otro espacio métrico). El plano con la métrica de "*Manhattan*" (en honor de la famosísima y bien cuadriculada ciudad) se define como $\mathbb{R}^2$ con la función de distancia

$$d_m((x_1,x_2),(y_1,y_2)) := |y_1 - x_1| + |y_2 - x_2|\,;$$

pensando que para ir de un punto a otro sólo se puede viajar en dirección horizontal o vertical (como en las calles de una ciudad).

a) Demuestra que la métrica de Manhattan cumple el Teorema 1.27.
b) Dibuja el conjunto de puntos con distancia 1 al origen; ¿cómo son los círculos?
c) ¿Cómo son los segmentos (definidos por la distancia como en el ejercicio anterior)?
d) ¿Cómo son los rayos?

1.11.2 Distancia de un punto a una recta



Consideremos el siguiente problema. Nos dan una recta

$$\ell : \quad \mathbf{n}\cdot\mathbf{x} = c$$

y un punto $\mathbf{p}$ (que puede estar fuera o dentro de ella); y se nos pregunta cuál es la distancia de $\mathbf{p}$ a ℓ, que podemos denotar con $d(\mathbf{p},\ell)$.

Es más fácil resolverlo si lo pensamos junto con un problema aparentemente más complicado: ¿cuál es el punto de ℓ más cercano a $\mathbf{p}$? Llamémoslo $\mathbf{q} \in \ell$, aunque sea incógnito. El meollo del asunto es que la recta que pasa por $\mathbf{p}$ y $\mathbf{q}$ debe ser ortogonal a ℓ. Es decir, si el segmento $\overline{\mathbf{pq}}$ es ortogonal a ℓ entonces $\mathbf{q}$ es el punto en ℓ más cercano a $\mathbf{p}$. Para demostrarlo considérese cualquier otro punto $\mathbf{x} \in \ell$ y el triángulo rectángulo $\mathbf{x}$, $\mathbf{q}$, $\mathbf{p}$ y apliquése el Teorema de Pitágoras para ver que $d(\mathbf{x},\mathbf{p}) \geq d(\mathbf{q},\mathbf{p})$. Como la dirección ortogonal a ℓ es precisamente $\mathbf{n}$, entonces debe existir $t \in \mathbb{R}$ tal que $\mathbf{q} - \mathbf{p} = t\,\mathbf{n}$. La nueva incógnita t nos servirá para medir la distancia de $\mathbf{p}$ a ℓ. Tenemos entonces que nuestras incógnitas $\mathbf{q}$ y t cumplen las siguientes ecuaciones:

$$\begin{aligned} \mathbf{n}\cdot\mathbf{q} &= c \\ \mathbf{q} - \mathbf{p} &= t\,\mathbf{n}, \end{aligned}$$

donde la primera dice que $\mathbf{q} \in \ell$ y la segunda que el segmento $\overline{\mathbf{pq}}$ es ortogonal a ℓ.

Tomando el producto interior de $\mathbf{n}$ con la segunda ecuación, obtenemos

$$\mathbf{n}\cdot\mathbf{q} - \mathbf{n}\cdot\mathbf{p} = t(\mathbf{n}\cdot\mathbf{n}).$$

De donde, al sustituir la primera ecuación, se puede despejar t (pues $\mathbf{n} \neq \mathbf{0}$) en puros términos conocidos:

$$\begin{aligned} c - (\mathbf{n} \cdot \mathbf{p}) &= t\,(\mathbf{n} \cdot \mathbf{n}) \\ t &= \frac{c - (\mathbf{n} \cdot \mathbf{p})}{\mathbf{n} \cdot \mathbf{n}}. \end{aligned}$$

De aquí se deduce que

$$\begin{aligned} \mathbf{q} &= \mathbf{p} + \left(\frac{c - (\mathbf{n} \cdot \mathbf{p})}{\mathbf{n} \cdot \mathbf{n}} \right) \mathbf{n} \\ d(\mathbf{p}, \ell) &= |t\,\mathbf{n}| = |t|\,|\mathbf{n}| \\ &= \frac{|c - (\mathbf{n} \cdot \mathbf{p})|}{|\mathbf{n} \cdot \mathbf{n}|}\,|\mathbf{n}| = \frac{|c - (\mathbf{n} \cdot \mathbf{p})|}{|\mathbf{n}|}. \end{aligned}$$

Puesto que sólo hemos usado las propiedades del producto interior y la norma, sin ninguna referencia a las coordenadas, podemos generalizar a $\mathbb{R}^3$; o bien a $\mathbb{R}^n$ pensando que un *hiperplano* (definido por una ecuación lineal $\mathbf{n} \cdot \mathbf{x} = c$) es una recta cuando $n = 2$ y es un plano cuando $n = 3$.

Proposicion 1.29 *Sea* Π *un hiperplano en* $\mathbb{R}^n$ *dado por la ecuación* $\mathbf{n} \cdot \mathbf{x} = c$, *y sea* $\mathbf{p} \in \mathbb{R}^n$ *cualquier punto. Entonces*

$$d(\mathbf{p}, \Pi) = \frac{|c - (\mathbf{n} \cdot \mathbf{p})|}{|\mathbf{n}|},$$

y la proyección ortogonal de $\mathbf{p}$ *sobre* Π *(es decir, el punto más cercano a* $\mathbf{p}$ *en* Π*) es*

$$\mathbf{q} = \mathbf{p} + \left(\frac{c - \mathbf{n} \cdot \mathbf{p}}{\mathbf{n} \cdot \mathbf{n}} \right) \mathbf{n}.$$

EJERCICIO 1.97 Evalúa las dos fórmulas de la Proposición 1.29 cuando $\mathbf{p} \in \Pi$.

EJERCICIO 1.98 Encuentra una expresión para $d(\mathbf{p}, \ell)$ cuando $\ell = \{\mathbf{q} + t\mathbf{v} \mid t \in \mathbb{R}\}$.

EJERCICIO 1.99 Demuestra que la fórmula de la distancia de un punto a un plano obtenida en el Ejercicio 1.89 coincide con la de la Proposición 1.29.

EJERCICIO 1.100 Encuentra la distancia del punto ... a la recta ... y su proyección ortogonal.

1.11.3 El determinante como área dirigida

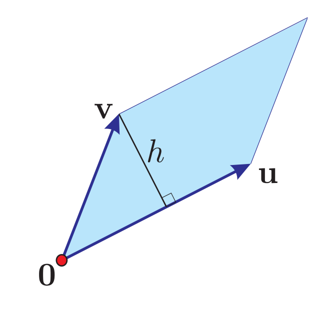
No estamos en posición de desarrollar una teoría general de la noción de área; para ello se requieren ideas de cálculo. Pero sí podemos trabajarla para figuras simples

como triángulos y paralelogramos. Se define el *área* de un paralelogramo de manera euclidiana (como en secundaria), es decir, $(\mathtt{base}) \times (\mathtt{altura})$. Veremos que se calcula con un viejo conocido.

Nos encontramos primero el determinante asociado a sistemas de dos ecuaciones lineales con dos incógnitas. Después, definimos el determinante de dos vectores en $\mathbb{R}^2$, otra vez por sistemas de ecuaciones y vimos que detecta (cuando es cero) el paralelismo. Ahora podemos darle una interpretación geométrica en todos los casos.

Teorema 1.30 *Sean* $\mathbf{u}$ *y* $\mathbf{v}$ *dos vectores cualesquiera en* $\mathbb{R}^2$*. El área del paralelogramo que definen* $\mathbf{u}$ *y* $\mathbf{v}$ *es* $|\det(\mathbf{u},\mathbf{v})|$*. Además* $\det(\mathbf{u},\mathbf{v}) \geq 0$ *cuando el movimiento angular más corto de* $\mathbf{u}$ *a* $\mathbf{v}$ *es positivo (* $\overrightarrow{\text{ang}}(\mathbf{u},\mathbf{v}) \in [0,\pi]$*) y* $\det(\mathbf{u},\mathbf{v}) \leq 0$ *cuando* $\overrightarrow{\text{ang}}(\mathbf{u},\mathbf{v}) \in [-\pi, 0]$.

Demostración. Podemos tomar como base del paralelogramo en cuestión al vector $\mathbf{u}$, es decir, $(\mathtt{base})$ en la fórmula $(\mathtt{base}) \times (\mathtt{altura})$ es $|\mathbf{u}|$, y suponemos que $\mathbf{u} \neq \mathbf{0}$. La altura, h digamos, es entonces la distancia de $\mathbf{v}$, pensado como punto, a la recta ℓ generada por $\mathbf{u}$. Esta recta tiene ecuación normal $\mathbf{u}^\perp \cdot \mathbf{x} = 0$. Entonces, por la Proposición 1.29 se tiene que

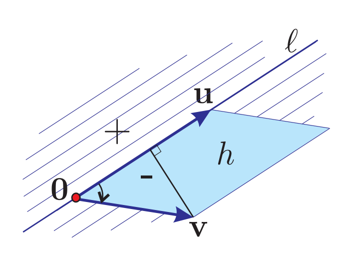


$$h = \frac{\left|0 - (\mathbf{u}^\perp \cdot \mathbf{v})\right|}{|\mathbf{u}^\perp|} = \frac{\left|\mathbf{u}^\perp \cdot \mathbf{v}\right|}{|\mathbf{u}|}.$$

Al multiplicar por la base se obtiene que el área del paralelogramo es el valor absoluto del determinante

$$\frac{\left|\mathbf{u}^\perp \cdot \mathbf{v}\right|}{|\mathbf{u}|}|\mathbf{u}| = \left|\mathbf{u}^\perp \cdot \mathbf{v}\right| = |\det(\mathbf{u},\mathbf{v})|\,.$$

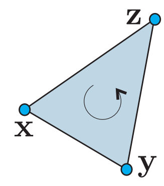
Para ver lo que significa el signo de $\det(\mathbf{u},\mathbf{v})$, obsérvese que la recta ℓ, generada por $\mathbf{u}$, parte el plano en dos *semiplanos*. En uno de ellos se encuentra su compadre ortogonal $\mathbf{u}^\perp$, éste es el lado *positivo*, pues al tomar $c \geq 0$ las ecuaciones $\mathbf{u}^\perp \cdot \mathbf{x} = c$ definen rectas paralelas a ℓ, y una de ellas (cuando $c = \left|\mathbf{u}^\perp\right|^2 > 0$) pasa por $\mathbf{u}^\perp$; como estas rectas llenan ("ashuran") todo ese semiplano, entonces $\mathbf{u}^\perp \cdot \mathbf{v} \geq 0$ si y sólo si $\mathbf{v}$ está en el lado positivo de ℓ. Análogamente, $\mathbf{u}^\perp \cdot \mathbf{v} < 0$ si y sólo si $\mathbf{v}$ está en el mismo lado de ℓ que $-\mathbf{u}^\perp$, que llamamos el lado *negativo.*



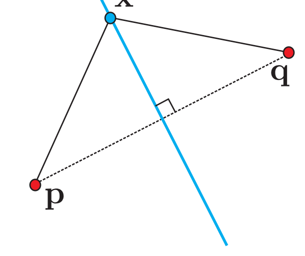
□

De esta manera, el determinante es un "área signada": no sólo nos da el área sino que también nos dice si los vectores están orientados positiva o negativamente. Obsérvese que nuestra manera de detectar el paralelismo corresponde entonces a que el área es cero (que el paralelogramo en cuestión es unidimensional).

EJERCICIO 1.101 El área signada de un triángulo *orientado* $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ se debe definir como $\det((\mathbf{y}-\mathbf{x}), (\mathbf{z}-\mathbf{x}))$. Demuestra que sólo depende del orden cíclico en que aparecen los vértices, es decir, que da lo mismo si tomamos como ternas a $\mathbf{y}$, $\mathbf{z}$, $\mathbf{x}$ o a $\mathbf{z}$, $\mathbf{x}$, $\mathbf{y}$. Pero que es su inverso aditivo si tomamos la otra orientación $\mathbf{z}$, $\mathbf{y}$, $\mathbf{x}$ para el triángulo.



1.11.4 La mediatriz



Cuando empecemos a estudiar las curvas cónicas, las definiremos como *lugares geométricos* de puntos que cumplen cierta propiedad que involucra distancias. Así que para terminar este capítulo veremos los ejemplos más sencillos.

Dados dos puntos $\mathbf{p}$ y $\mathbf{q}$ ($\in \mathbb{R}^2$), ¿cuál es el lugar geométrico de los puntos que equidistan de ellos? Es decir, hay que describir al conjunto

$$\left\{\mathbf{x} \in \mathbb{R}^2 \mid d(\mathbf{p},\mathbf{x}) = d(\mathbf{q},\mathbf{x})\right\}.$$

Al desarrollar la ecuación $d(\mathbf{p},\mathbf{x})^2 = d(\mathbf{q},\mathbf{x})^2$, que equivale a la anterior pues las distancias son positivas, tenemos

$$\begin{aligned}
(\mathbf{p}-\mathbf{x})\cdot(\mathbf{p}-\mathbf{x}) &= (\mathbf{q}-\mathbf{x})\cdot(\mathbf{q}-\mathbf{x}) \\
\mathbf{p}\cdot\mathbf{p} - 2(\mathbf{p}\cdot\mathbf{x}) + \mathbf{x}\cdot\mathbf{x} &= \mathbf{q}\cdot\mathbf{q} - 2(\mathbf{q}\cdot\mathbf{x}) + \mathbf{x}\cdot\mathbf{x} \\
\mathbf{p}\cdot\mathbf{p} - \mathbf{q}\cdot\mathbf{q} &= 2(\mathbf{p}-\mathbf{q})\cdot\mathbf{x}
\end{aligned}$$

que se puede reescribir como

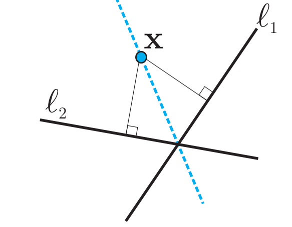
$$(\mathbf{p}-\mathbf{q})\cdot\mathbf{x} = (\mathbf{p}-\mathbf{q})\cdot(\frac{1}{2}(\mathbf{p}+\mathbf{q})).$$

Y ésta es la ecuación normal de la recta ortogonal al segmento $\overline{\mathbf{pq}}$ y que lo intersecta en su punto medio, llamada la *mediatriz* de $\mathbf{p}$ y $\mathbf{q}$.

EJERCICIO 1.102 ¿Cuál es el lugar geométrico de los puntos en $\mathbb{R}^3$ que equidistan de dos puntos? Describe la demostración.

1.11.5 Bisectrices y ecuaciones unitarias

Preguntamos ahora cuál es el lugar geométrico de los puntos que equidistan de dos rectas ℓ_1 y ℓ_2. De todas las ecuaciones normales que definen a estas rectas, conviene escoger una donde el vector normal es unitario para que las distancias sean más fáciles de expresar. Tal ecuación se llama *unitaria* y se obtiene de cualquier ecuación normal dividiendo —ambos lados de la ecuación, por supuesto— entre la norma del vector normal. Sean entonces $\mathbf{u}_1$, $\mathbf{u}_2 \in \mathbb{S}^1$ y $c_1, c_2 \in \mathbb{R}$ tales que para $i = 1, 2$



$$\ell_i : \mathbf{u}_i \cdot \mathbf{x} = c_i.$$

Por la fórmula de la sección 1.11.2 se tiene entonces que los puntos $\mathbf{x}$ que equidistan de ℓ_1 y ℓ_2 son precisamente los que cumplen con la ecuación

$$|\mathbf{u}_1 \cdot \mathbf{x} - c_1| = |\mathbf{u}_2 \cdot \mathbf{x} - c_2| .$$

Ahora bien, si el valor absoluto de dos números reales coincide entonces o son iguales o son inversos aditivos, lo cual se expresa elegantemente de la siguiente manera:

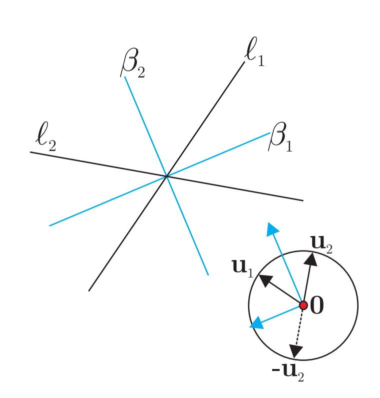
$$|a| = |b| \quad \Leftrightarrow \quad (a+b)(a-b) = 0,$$

que, aplicado a nuestra ecuación anterior, nos dice que el lugar geométrico que estamos buscando está determinado por la ecuación

$$\begin{aligned} (\mathbf{u}_1 \cdot \mathbf{x} - c_1 + \mathbf{u}_2 \cdot \mathbf{x} - c_2)(\mathbf{u}_1 \cdot \mathbf{x} - c_1 - \mathbf{u}_2 \cdot \mathbf{x} + c_2) &= 0 \\ ((\mathbf{u}_1 + \mathbf{u}_2) \cdot \mathbf{x} - (c_1 + c_2))((\mathbf{u}_1 - \mathbf{u}_2) \cdot \mathbf{x} - (c_1 - c_2)) &= 0. \end{aligned}$$

Como cuando el producto de dos números es cero alguno de ellos lo es, entonces nuestro lugar geométrico es la unión de las dos rectas:

$$\begin{aligned} \beta_1 &: \quad (\mathbf{u}_1 + \mathbf{u}_2) \cdot \mathbf{x} = (c_1 + c_2) \qquad (1.12) \\ \beta_2 &: \quad (\mathbf{u}_1 - \mathbf{u}_2) \cdot \mathbf{x} = (c_1 - c_2), \end{aligned}$$



que son las *bisectrices* de las lineas ℓ_1 y ℓ_2. Observemos que son ortogonales pues

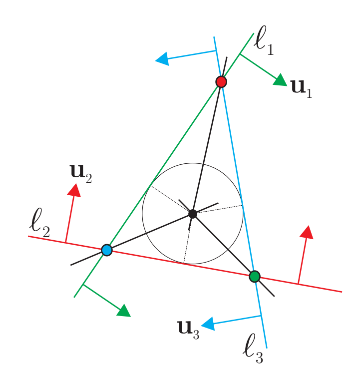
$$\begin{aligned} (\mathbf{u}_1 + \mathbf{u}_2) \cdot (\mathbf{u}_1 - \mathbf{u}_2) &= |\mathbf{u}_1|^2 - |\mathbf{u}_2|^2 \\ &= 1 - 1 \\ &= 0, \end{aligned}$$

y, geométricamente, bisectan a los dos sectores o ángulos que forman las rectas; también pasan por el punto de intersección pues sus ecuaciones se obtienen sumando y restando las ecuaciones originales (el caso en que sean paralelas se deja como ejercicio). En el círculo unitario tambien es claro que la suma y la diferencia de dos vectores (aunque ya no sean necesariamente unitarios) tienen los ángulos adecuados.

Como corolario, obtenemos otro de los teoremas clásicos de concurrencia.

Teorema 1.31 *Las bisectrices (internas) de un triángulo son concurrentes.*

Demostración. Sólo hay que tener cuidado con cuáles ecuaciones unitarias trabajamos, pues hay dos posibles para cada recta. Si escogemos los vectores unitarios de las tres rectas, ℓ_1, ℓ_2 y ℓ_3 digamos, apuntando hacia adentro del triángulo, $\mathbf{u}_1$, $\mathbf{u}_2$ y $\mathbf{u}_3$ respectivamente, entonces está claro que sus sumas son normales a las bisectrices exteriores. El teorema habla entonces de las bisectrices que se obtienen como diferencias, que son, numerándolas por el índice de la recta opuesta:



$$\beta_1 \quad : \quad (\mathbf{u}_2 - \mathbf{u}_3) \cdot \mathbf{x} = (c_2 - c_3)$$
$$\beta_2 \quad : \quad (\mathbf{u}_3 - \mathbf{u}_1) \cdot \mathbf{x} = (c_3 - c_1)$$
$$\beta_3 \quad : \quad (\mathbf{u}_1 - \mathbf{u}_2) \cdot \mathbf{x} = (c_1 - c_2) .$$

El negativo de cualquiera de ellas se obtiene sumando las otras dos.

□

EJERCICIO 1.103 Un argumento aparentemente más elemental para el teorema anterior es que si un punto está en la intersección de dos bisectrices, entonces ya equidista de las tres rectas y por tanto también está en la bisectriz del vertice restante. ¿Qué implica este argumento cuando tomamos bisectrices exteriores? ¿En dónde concurren dos bisectrices exteriores? ¿Una interior y una exterior? Da enunciados y demuéstralos con ecuaciones normales. Completa el dibujo.

EJERCICIO 1.104 Demuestra que el lugar geométrico de los puntos que equidistan a dos rectas paralelas es la paralela a ambas que pasa por el punto medio de sus intersecciones con una recta no paralela a ellas, a partir de las ecuaciones (1.12).

EJERCICIO 1.105 ¿Cuál es el lugar geométrico de los puntos que equidistan de dos planos en $\mathbb{R}^3$? Describe una demostración de tu aseveración.

1.12 *Los espacios de rectas en el plano

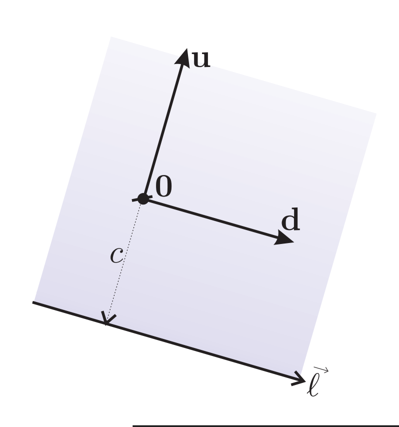
Esta sección "asterisco" o extra es, hasta cierto punto, independiente del texto principal; será relevante como motivación hasta el capítulo 6, donde se estudia el plano proyectivo y cristaliza en la llamada "dualidad". Estudiaremos dos ejemplos que ya están a la mano de cómo la noción de *el Espacio*, que por milenios se consideró única y amorfa —"el espacio en que vivimos"—, se pluraliza *los* espacios. Ya vimos que el Espacio teórico que consideraron los griegos es sólo la instancia en dimensión 3

de la familia de espacios euclidianos $\mathbb{E}^n$; pero además la *pluralización* (pasar de uno a muchos) se da en otro sentido. En matemáticas, y en particular en la geometría, surgen naturalmente otros conjuntos, además de los $\mathbb{E}^n$, con plenos derechos de ser llamados "espacios"; en la física se les llama "espacios fase". Los dos "espacios"[5] que nos ocupan ahora son los de rectas en el plano. Consideraremos cada recta como un punto de un nuevo "espacio" y, puesto que intuitivamente podemos decir si dos rectas son cercanas o no, en este nuevo "espacio", sabremos si dos puntos lo están. La idea clave está en que así como Descartes asoció parejas de números a los puntos del plano, a las rectas del plano también hemos asociado números (los involucrados en sus ecuaciones) de tal manera que si esos números varían un poco la recta asociada también se mueve poco; dan lugar a una noción de "continuidad" que es la noción básica de la topología.

En el Ejercicio 1.68 de la sección 1.8, se hizo notar que a cada terna de números $(a, b, c) \in \mathbb{R}^3$, con $(a, b) \neq (0, 0)$, se le puede asociar una recta en $\mathbb{R}^2$, a saber, la definida por la ecuación $ax + by = c$; el estudiante debió haber demostrado que las rectas por el origen en $\mathbb{R}^3$ dan la misma recta en $\mathbb{R}^2$ y que los planos por el origen en $\mathbb{R}^3$ corresponden a los haces de rectas en el plano. Dando esto por entendido, en esta última sección abundaremos en tales ideas. Se puede decir que la pregunta que nos guía es: ¿cuántas rectas hay en $\mathbb{R}^2$?, pero el término "cuántas" es demasiado impreciso. Veremos que las rectas de $\mathbb{R}^2$ forman algo que llamaremos el "espacio" de rectas (un caso especial de las llamadas "Grassmannianas" pues quien primero desarrollo estas ideas fue Grassmann), y para esto les daremos "coordenadas" explícitas, muy a semejanza de las coordenadas polares para los puntos.

1.12.1 Rectas orientadas

Consideremos primero las rectas orientadas en $\mathbb{R}^2$, pues a ellas se les puede asociar un vector normal (y por tanto una ecuación) de manera canónica. Una *recta orientada* es una recta $\vec{\ell}$ junto con una orientación preferida, que nos dice la manera *positiva* de viajar en ella. Entonces, de todos sus vectores direccionales podemos escoger el unitario $\mathbf{d}$ con la orientación positiva y, como vector normal, escogemos el compadre ortogonal de éste; es decir, el vector unitario tal que después de la orientación de la recta nos da la orientación positiva del plano. Así, cada recta orientada está definida por una ecuación *unitaria* única

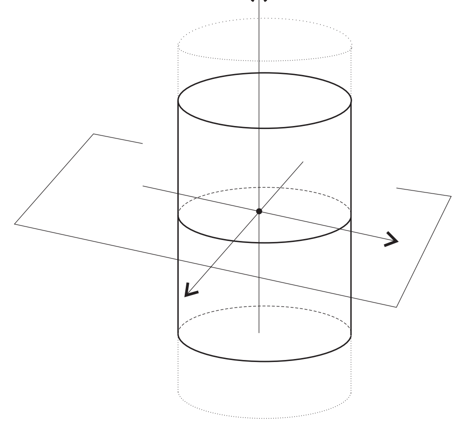


$$\mathbf{u} \cdot \mathbf{x} = c,$$

[5] Mantenemos las comillas pues la noción aún no está bien definida: va en la dirección de lo que ahora se llama topología.

donde $\mathbf{u} \in \mathbb{S}^1$, y $c \in \mathbb{R}$ es su distancia orientada al origen.

Nótese que a cada recta orientada se le puede asociar su semiplano positivo $\mathbf{u} \cdot \mathbf{x} \geq c$; así que también podemos pensar a las rectas orientadas como semiplanos.



Puesto que esta ecuación unitaria es única (por la orientación) entonces hay tantas rectas orientadas como parejas

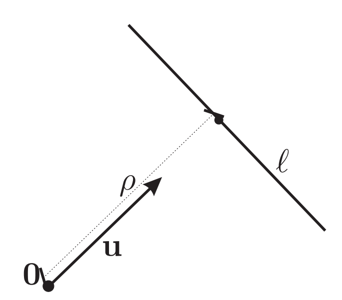
$$(\mathbf{u}, c) \in \mathbb{S}^1 \times \mathbb{R} \subset \mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3,$$

y este conjunto es claramente un cilindro vértical en $\mathbb{R}^3$. Los haces de líneas orientadas paralelas corresponden a líneas verticales en el cilindro, pues en estos haces el vector normal está fijo. Y los haces de líneas (orientadas) concurrentes corresponden a lo que corta un plano por el origen al cilindro $\mathbb{S}^1 \times \mathbb{R}$ (que, como veremos en el capítulo siguiente, se llama elípse).

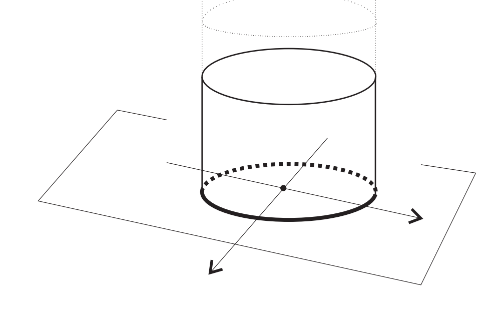
EJERCICIO 1.106 Demuestra la última afirmación.

1.12.2 Rectas no orientadas

Las dos orientaciones de una recta ℓ tienen ecuaciones unitarias $\mathbf{u} \cdot \mathbf{x} = c$ y $(-\mathbf{u}) \cdot \mathbf{x} = -c$. Si la recta no pasa por el origen (si $0 \notin \ell$ y por tanto $c \neq 0$), de estas dos ecuaciones unitarias podemos tomar la ecuación con constante estrictamente positiva, $\mathbf{u} \cdot \mathbf{x} = \rho$, con $\rho > 0$, digamos. Así que las "coordenadas" $(\mathbf{u}, \rho)$ corresponden a las coordenadas polares del punto en la recta ℓ más cercano al origen; y viceversa, a cada punto $\mathbf{x} \in \mathbb{R}^2 - \{0\}$ se le asocia la recta que pasa por el punto $\mathbf{x}$ y es normal al vector $\mathbf{x}$. O dicho de otra manera, si la recta ℓ no contiene al origen, podemos escoger el semiplano que define que no contiene a 0. Nos queda una ambigüedad en el origen: para las rectas que pasan por él tenemos dos vectores unitarios normales que definen la misma recta. Podemos resumir que las rectas en $\mathbb{R}^2$ corresponden a los puntos de un cilindro "semicerrado"

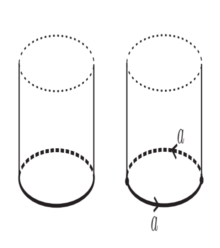


$$\mathbb{S}^1 \times \mathbb{R}_+,$$

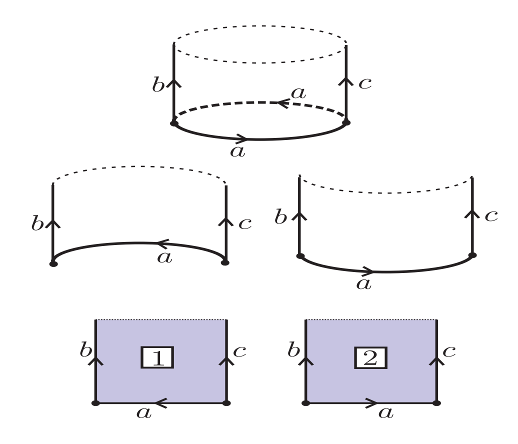


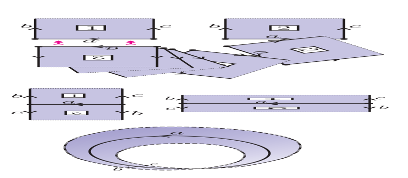
(donde $\mathbb{R}_+ = \{\rho \in \mathbb{R} \mid \rho \geq 0\}$), donde hay que identificar su frontera por antípodas, es decir, en este conjunto $(\mathbf{u}, 0)$ y $(-\mathbf{u}, 0)$ aún representan la misma recta. Si identificamos $(\mathbf{u}, 0)$ con $(-\mathbf{u}, 0)$ en el cilindro $\mathbb{S}^1 \times \mathbb{R}_+$, se obtiene una banda de Moebius abierta. Veámos:

Hay que pensar que el cilindro $\mathbb{S}^1 \times \mathbb{R}_+$ está hecho de un material flexible (pensarlo como "espacio topológico", se dice actualmente), y hacer muy pequeño el factor $\mathbb{R}_+$, para tener un cilindrito con una de sus fronteras inexistente (la que corresponde al infinito), indicada en las figuras con línea punteada. En la otra frontera (que corresponde al $\mathsf{0}$) debemos pegar puntos antípodas. Llamemos a al segmento dirigido que es la mitad de ese círculo y entonces la otra mitad, con la misma orientación del círculo, vuelve a ser a.



Si tratamos de pegar directamente los segmentos a, la superficie del cilindro nos estorba y no se ve qué pasa con claridad (de hecho no se puede lograr en $\mathbb{R}^3$ sin que la superficie se autointersecte). Conviene entonces cortar el cilindro por los segmentos verticales, denotados con b y c en la figura, sobre el principio y fin de a, para obtener dos pequeños cuadrados (que hemos bautizado 1 y 2). Los segmentos b y c están dirigidos y debemos recordar que hay que volverlos a pegar cuando se pueda.





Ahora sí podemos pegar los segmentos a recordando su dirección. Basta voltear el cuadrado 2 de cabeza para lograrlo, y obtener un rectángulo. Si alargamos el rectángulo en su dirección horizontal, obtenemos una banda en la cual hay que identificar las fronteras verticales hechas, cada una, con los pares de segmentos b y c (la frontera horizontal sigue siendo punteada). Y esta identificación, según indican nuestros segmentos dirigidos, es invirtiendo la orientación; hay que dar media vuelta para pegarlos, así que el resultado es una banda de Moebius a la que se le quita su frontera, y ya no queda nada por identificar.

Hemos demostrado que las rectas del plano corresponden biyectivamente a los puntos de una banda de Moebius abierta, y que esta correspondencia preserva "cercanía" o "continuidad".

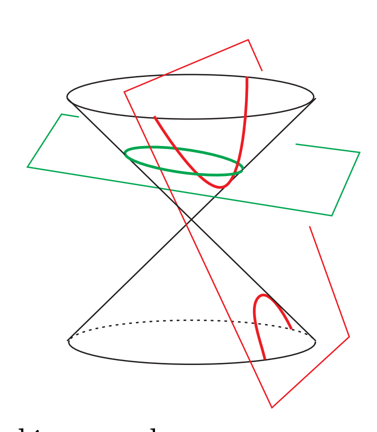
Obsérvese además que las rectas que pasan por el origen (el segmento $\mathfrak{a}$ con sus extremos identificados) quedan en el "corazón" (el círculo central) de la banda de Moebius; los haces paralelos corresponden a los intervalos abiertos transversales al corazón que van de la frontera a sí misma (el punto de intersección con el corazón es su representante en el origen). ¿A qué corresponden los haces concurrentes?

EJERCICIO 1.107 Demuestra que dos haces de rectas, no ambos haces paralelos, tienen una única recta en comun.

Capítulo 2

Cónicas I (presentación)

Las cónicas son una familia de curvas famosas que definieron los griegos. De ellas, la que vemos cientos de veces al día es la elipse. Al ver la boca de un vaso, una taza o una vasija, o bien, la llanta de un coche, vemos una *elipse* e inmediatamente intuimos que en realidad se trata de un círculo, sólo que lo estamos viendo de lado. Únicamente vemos un círculo cuando la proyección ortogonal de nuestro ojo al plano en el que vive el círculo es justo su centro; así que los cientos de círculos que vemos a diario son nuestra deducción cerebral automática de las elipses que en realidad percibimos. Para fijar ideas, pensemos que vemos con un solo ojo una taza sobre una mesa; cada punto de su borde define una recta a nuestro ojo por donde viaja la luz que percibimos; todas estas rectas forman un *cono* con *ápice* $\times$en el punto ideal de nuestro ojo; pero no es un cono circular sino un cono elíptico: más apachurrado entre más cerca esté nuestro ojo al plano de la boca de la taza, y más circular en cuanto nuestro ojo se acerca a verlo desde arriba. Resulta entonces que si a este cono elíptico lo cortamos con el plano de la boca de la taza nos da un círculo (precisamente, la boca de la taza). Los griegos intuyeron entonces que, al revés, la elipse se puede definir cortando conos circulares (definidos con base en el círculo que ya conocían) con planos inclinados.



Hoy día lo podemos decir de otra manera: si prendemos una lámpara de mano, de ella emana un cono circular de luz; es decir, si la apuntamos ortogonalmente a una pared se ilumina un círculo (con todo y su interior); al girar un poco la lámpara, el círculo se deforma en una elipse que es un plano (la pared) que corta un cono circular (el haz de luz). Entre más inclinada esté la lámpara respecto a la pared, la elipse se alarga (como boca de taza con el ojo casi al ras) hasta que algo de la luz sale de la pared. Pensemos entonces que la pared es infinita (un plano) y que nuestra lámpara es infinitamente potente (un cono perfecto); al girar más la lámpara, la elipse se sigue alargando y alargando hasta el momento en que uno de los rayos del cono no toca la pared (es paralelo a ella). Justo en ese instante, el borde de lo iluminado en la pared es

una *parábola*, y de allí en adelante, al continuar el giro, es una *rama de hipérbola*. La *hipérbola* completa se obtiene al pensar que los rayos del cono son rectas completas, que se continúan hacia atrás de la lámpara, y entonces algunas de estas rectas (cuyos rayos de luz ya no tocan la pared por adelante) la intersectan por atrás en la otra rama de la hipérbola. Si continuamos el giro, pasamos de nuevo por una parábola y luego por elipses que, viniendo de atrás, se redondean hasta el círculo con el que empezamos, pero ahora dibujado por el haz de rayos "hacia atrás" de la lámpara que giró $180°$. La definición clásica de los griegos, que es equivalente, es fijando el cono y moviendo el plano, y de allí que las hayan llamado *secciones cónicas* o simplemente *cónicas*: la intersección de planos con conos circulares (completos).

Los geómetras griegos estudiaron con profundidad estas curvas. Lo hicieron por amor al arte, al conocimiento abstracto y puro, por la intuición matemática de su propia belleza y naturalidad; y dieron a la humanidad una lección fundamental de la importancia que tiene la ciencia básica. Pues resultó, casi dos mil años después, que estas curvas teóricas se expresaban en la naturaleza como las órbitas de los planetas (elipses descritas por Kepler) o en las trayectorias de los proyectiles (parábolas descritas por Newton y predecesores). Sus propiedades focales, que también descubrieron los griegos, se usan hoy cotidianamente en las antenas parabólicas, en los telescopios y en el diseño de reflectores o de las lentes de anteojos o cámaras. Su teoría, que parecía ser un divertimento totalmente abstracto, resultó importantísima en la vida diaria y en el entendimiento de la naturaleza.

Los geómetras griegos demostraron que a las cónicas también se les puede definir intrínsecamente como *lugares geométricos* en el plano, es decir, como subconjuntos de puntos que cumplen cierta propiedad; a saber, una propiedad que se expresa en términos de distancias. Ésta será la definición formal que usaremos en el libro y en la última sección de este capítulo veremos que coincide con la de secciones cónicas. En las primeras secciones veremos que su definición como lugares geométricos expresados en términos de distancias es equivalente a su descripción como el lugar geométrico de puntos en $\mathbb{R}^2$ cuyas coordenadas satisfacen cierta ecuación. Resulta que estas ecuaciones son ecuaciones cuadráticas en dos variables, x y y (las coordenadas). Este hecho, que observó el propio Descartes, fue una gran motivación para el surgimiento de su geometría analítica, pues relacionaba las clásicas secciones cónicas con los polinomios de segundo grado, y en general el álgebra, que habían desarrollado los árabes.

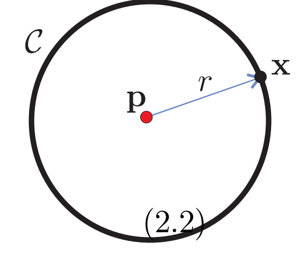
2.1 Círculos

Ya hemos definido y usado extensamente el círculo unitario $\mathbb{S}^1$, determinado por la ecuación

$$x^2 + y^2 = 1.$$

Consideremos ahora cualquier otro círculo $\mathcal{C}$. Tiene un *centro* $\mathbf{p} = (h, k)$, un *radio* $r > 0$, y es el lugar geométrico de los puntos cuya distancia a $\mathbf{p}$ es r. Es decir, $\mathcal{C} = \left\{\mathbf{x} \in \mathbb{R}^2 \mid d(\mathbf{x}, \mathbf{p}) = r\right\}$; o bien, $\mathcal{C}$ está definido por la ecuación

$$d(\mathbf{x}, \mathbf{p}) = r. \tag{2.1}$$



Puesto que ambos lados de esta ecuación son positivos, es equivalente a la igualdad de sus cuadrados que en coordenadas toma la forma

$$(x - h)^2 + (y - k)^2 = r^2. \tag{2.2}$$

De tal manera que todos los círculos de $\mathbb{R}^2$ están determinados por una ecuación cuadrática en las variables x y y. Cuando la ecuación tiene la forma anterior, podemos "leer" toda la información geométrica (el centro y el radio). Pero en general viene disfrazada como

$$x^2 + y^2 - 2hx - 2ky = \left(r^2 - h^2 - k^2\right)$$

que es, claramente, su expresión desarrollada. Veamos un ejemplo.

Consideremos la ecuación

$$x^2 + y^2 - 6x + 2y = -6. \tag{2.3}$$

Para ver si define un círculo, y cuál, hay que tratar de escribirla en la forma (2.2). Para esto hay que "completar" los cuadrados sumando (en ambos lados) las constantes que faltan

$$\begin{aligned} \left(x^2 - 6x + 9\right) + \left(y^2 + 2y + 1\right) &= -6 + 9 + 1 \\ (x - 3)^2 + (y + 1)^2 &= 4. \end{aligned}$$

Es claro que al desarrollar esta ecuación obtenemos la original, de tal manera que aquélla define el círculo con centro en $(3, -1)$ y radio 2.

También podemos expresar la ecuación de un círculo de manera vectorial, pues el círculo $\mathcal{C}$ con centro $\mathbf{p}$ y radio r está claramente definido por la ecuación

$$(\mathbf{x} - \mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}) = r^2, \tag{2.4}$$

que es el cuadrado de (2.1). Esta ecuación, que llamaremos la *ecuación vectorial del círculo,* se puede también reescribir como

$$\mathbf{x} \cdot \mathbf{x} - 2\mathbf{p} \cdot \mathbf{x} + \mathbf{p} \cdot \mathbf{p} = r^2.$$

Parece superfluo pero tiene grandes ventajas. Por ejemplo, en algunos de los ejercicios siguientes puede ayudar, pero además, y mucho más profundo, al no hacer referencia a las coordenadas tiene sentido en cualquier dimensión. Sirve entonces para definir esferas en $\mathbb{R}^3$ y, como veremos en la siguiente sección (en la que vale la pena

que el estudiante tenga en mente cómo traducir a $\mathbb{R}^3$, esferas y planos), de la ecuación vectorial se puede extraer mucha información geométrica interesante.

Debemos también remarcar que en las siguientes secciones, donde empezamos a estudiar las otras cónicas, no aparece el equivalente de esta ecuación vectorial del círculo. Para lograrlo en general (en el capítulo 9) nos falta desarrollar mucha más teoría.

EJERCICIO 2.1 ¿Cuáles de las siguientes ecuaciones son ecuaciones de un círculo? Y en su caso, ¿de cuál?
a) $x^2 - 6x + y^2 - 4y = 12$
b) $x^2 + 4x + y^2 + 2y = 11$
c) $2x^2 + 8x + 2y^2 - 4y = -8$
d) $x^2 - 4x + y^2 - 2y = -6$
e) $4x^2 + 4x + y^2 - 2y = 4$.

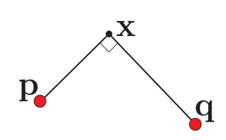
EJERCICIO 2.2 ¿Cuál es el lugar geométrico de los centros de los círculos que pasan por dos puntos (distintos) $\mathbf{a}$ y $\mathbf{b}$?

EJERCICIO 2.3 Sean $\mathbf{a}$ y $\mathbf{b}$ dos puntos (distintos) en el plano; y sean $\mathcal{C}_1$ el círculo con centro $\mathbf{a}$ y radio r_1 y $\mathcal{C}_2$ el círculo con centro $\mathbf{b}$ y radio r_2. Demuestra que $\mathcal{C}_1$ y $\mathcal{C}_2$ son *tangentes* (i.e. se intersectan en un único punto) si y sólo si $r_1 + r_2 = d(\mathbf{a}, \mathbf{b})$ o $r_1 + d(\mathbf{a}, \mathbf{b}) = r_2$ o $r_2 + d(\mathbf{a}, \mathbf{b}) = r_1$. ¿A qué corresponden geométricamente las tres condiciones anteriores? Haz un dibujo de cada caso.

EJERCICIO 2.4 Dados los círculos del ejercicio anterior, da condiciones sobre sus radios y la distancia entre sus centros para que:
a) el círculo $\mathcal{C}_1$ esté contenido en el interior del círculo $\mathcal{C}_2$;
b) el círculo $\mathcal{C}_1$ esté contenido en el exterior del círculo $\mathcal{C}_2$ (y viceversa);
c) el círculo $\mathcal{C}_1$ y el círculo $\mathcal{C}_2$ se intersecten en dos puntos.

EJERCICIO 2.5 Demuestra que por tres puntos no colineales $\mathbf{a}$,$\mathbf{b}$ y $\mathbf{c}$ pasa un único círculo, llamado el *circuncírculo* del triángulo $\mathbf{a}$,$\mathbf{b}$,$\mathbf{c}$.

EJERCICIO 2.6 Demuestra que dadas tres líneas no concurrentes y no paralelas dos a dos existen exactamente cuatro círculos tangentes a las tres; al que está contenido en el interior del triángulo se le llama su *incírculo*.



EJERCICIO 2.7 Sean $\mathbf{p}$ y $\mathbf{q}$ dos puntos distintos en el plano (o en $\mathbb{R}^3$). Demuestra que el conjunto de puntos $\mathbf{x}$ cuyas líneas a $\mathbf{p}$ y a $\mathbf{q}$ son ortogonales, es un círculo (o una esfera) que "quiere contener" a $\mathbf{p}$ y a $\mathbf{q}$. Más precisamente, que el conjunto definido por la ecuación

$$(\mathbf{x} - \mathbf{p}) \cdot (\mathbf{x} - \mathbf{q}) = 0$$

es el círculo con el segmento $\overline{\mathbf{pq}}$ como diametro. *(Compara esto con el ejercicio siguiente).*

EJERCICIO 2.8 Sean $\mathbf{p}$ y $\mathbf{q}$ dos puntos distintos en el plano. ¿Para cuáles números reales c se tiene que la ecuación

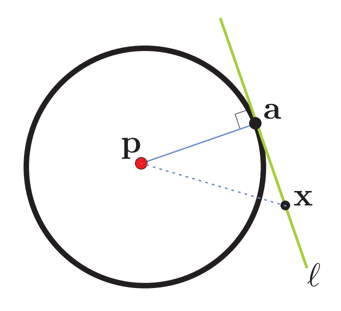
$$(\mathbf{x}-\mathbf{p})\cdot(\mathbf{x}-\mathbf{q})=c$$

define un círculo? En su caso, ¿cuál es el radio y dónde está su centro?

2.1.1 Tangentes y polares

Observemos primero que las líneas *tangentes* a un círculo son las normales a los *radios* (los segmentos del centro a sus puntos). Efectivamente, si $\mathbf{a}$ es un punto del círculo $\mathcal{C}$ dado por la ecuación vectorial

$$(\mathbf{x}-\mathbf{p})\cdot(\mathbf{x}-\mathbf{p})=r^2, \tag{2.4}$$



entonces su tangente es la recta ℓ normal a $(\mathbf{a}-\mathbf{p})$ y que pasa por $\mathbf{a}$. Pues $\mathbf{a}$ es el punto más cercano a $\mathbf{p}$ en esta recta, de tal manera que para cualquier otro punto $\mathbf{x}\in\ell$ se tiene que $d(\mathbf{x},\mathbf{p})>r$. Como el círculo $\mathcal{C}$ parte el plano en dos pedazos (el *interior* donde $d(\mathbf{x},\mathbf{p})<r$, y el *exterior* donde $d(\mathbf{x},\mathbf{p})>r$), entonces ℓ está contenida en el exterior salvo por el punto $\mathbf{a}\in\mathcal{C}$; ésta será nuestra definición de *tangente*.

Claramente, la recta ℓ está dada por la ecuación

$$\mathbf{x}\cdot(\mathbf{a}-\mathbf{p})=\mathbf{a}\cdot(\mathbf{a}-\mathbf{p}),$$

que tiene una manera mucho más interesante de escribirse. Restando $\mathbf{p}\cdot(\mathbf{a}-\mathbf{p})$ a ambos lados se obtiene

$$\begin{aligned}\mathbf{x}\cdot(\mathbf{a}-\mathbf{p})-\mathbf{p}\cdot(\mathbf{a}-\mathbf{p}) &= \mathbf{a}\cdot(\mathbf{a}-\mathbf{p})-\mathbf{p}\cdot(\mathbf{a}-\mathbf{p})\\ (\mathbf{x}-\mathbf{p})\cdot(\mathbf{a}-\mathbf{p}) &= (\mathbf{a}-\mathbf{p})\cdot(\mathbf{a}-\mathbf{p})\\ (\mathbf{x}-\mathbf{p})\cdot(\mathbf{a}-\mathbf{p}) &= r^2,\end{aligned}$$

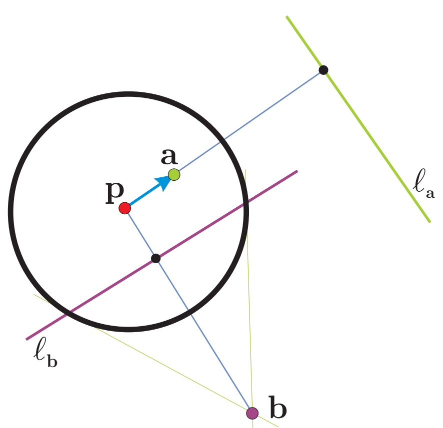
pues $\mathbf{a}\in\mathcal{C}$. La ecuación de la tangente se obtiene entonces al sustituir el punto en una de las dos apariciones de la variable $\mathbf{x}$ en la ecuación vectorial (2.4).

Para cualquier otro punto en el plano, $\mathbf{a}$ digamos, que no sea el centro ($\mathbf{a}\neq\mathbf{p}$), el mismo proceso algebraico —sustituir $\mathbf{a}$ en una de las instancias de $\mathbf{x}$ en la ecuación vectorial— nos da la ecuación de una recta, llamada *la polar* de $\mathbf{a}$ respecto al círculo $\mathcal{C}$, y que denotaremos $\ell_{\mathbf{a}}$; es decir,

$$\text{sea}\quad \ell_{\mathbf{a}}:\quad (\mathbf{x}-\mathbf{p})\cdot(\mathbf{a}-\mathbf{p})=r^2. \tag{2.5}$$

Ya hemos visto que cuando $\mathbf{a}\in\mathcal{C}$, su polar $\ell_{\mathbf{a}}$ es su tangente. Ahora veremos que si $\mathbf{a}$ está en el interior del círculo, entonces $\ell_{\mathbf{a}}$ no lo intersecta (está totalmente

contenida en el exterior), y que si está en el exterior (el punto $\mathbf{b}$ en la figura siguiente), entonces lo corta, y además lo corta en los dos puntos de $\mathcal{C}$ a los cuales se pueden trazar tangentes.



Para esto, expresemos la ecuación de $\ell_{\mathbf{a}}$ en su forma normal; desarrollando (2.5) se obtiene:

$$\mathbf{x} \cdot (\mathbf{a} - \mathbf{p}) = r^2 + \mathbf{p} \cdot (\mathbf{a} - \mathbf{p}) .$$

Esto indica que $\ell_{\mathbf{a}}$ es perpendicular al vector que va de $\mathbf{p}$ a $\mathbf{a}$. Ahora veamos cuál es su punto de intersección con la recta que pasa por $\mathbf{p}$ y $\mathbf{a}$. Parametricemos esta última recta con $\mathbf{p}$ de cero y $\mathbf{a}$ de uno (es decir como $\mathbf{p} + t(\mathbf{a} - \mathbf{p})$) y podemos despejar t al sustituir en la variable $\mathbf{x}$ de la ecuación anterior (o bien, esto se ve más directo al sustituir en la (2.5)), para obtener

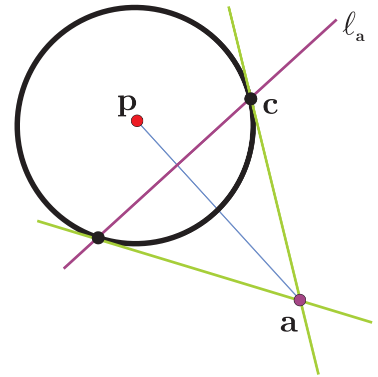
$$t = \frac{r^2}{(\mathbf{a} - \mathbf{p}) \cdot (\mathbf{a} - \mathbf{p})} = \frac{r^2}{d(\mathbf{p}, \mathbf{a})^2} .$$

Entonces la distancia de $\mathbf{p}$ a $\ell_{\mathbf{a}}$ es

$$d(\mathbf{p}, \ell_{\mathbf{a}}) = t\, d(\mathbf{p}, \mathbf{a}) = \frac{r^2}{d(\mathbf{p}, \mathbf{a})} = \left(\frac{r}{d(\mathbf{p}, \mathbf{a})} \right) r \tag{2.6}$$

y tenemos lo primero que queríamos: si $d(\mathbf{p}, \mathbf{a}) < r$ entonces $d(\mathbf{p}, \ell_{\mathbf{a}}) > r$; y al revés, si $d(\mathbf{p}, \mathbf{a}) > r$ entonces $d(\mathbf{p}, \ell_{\mathbf{a}}) < r$. Si el punto $\mathbf{a}$ está muy cerca de $\mathbf{p}$, su polar está muy lejos, y al revés, sus distancias al centro $\mathbf{p}$ se comportan como inversos pero "alrededor de r" (si $r = 1$ son precisamente inversos).

Supongamos ahora que $d(\mathbf{p}, \mathbf{a}) > r$, y sea $\mathbf{c}$ un punto en $\ell_{\mathbf{a}} \cap \mathcal{C}$ (que sabemos que existe pues $\ell_{\mathbf{a}}$ pasa por el interior de $\mathcal{C}$). Puesto que $\mathbf{c} \in \ell_{\mathbf{a}}$, se cumple la ecuación



$$(\mathbf{c} - \mathbf{p}) \cdot (\mathbf{a} - \mathbf{p}) = r^2 .$$

Pero entonces $\mathbf{a}$ cumple la ecuación de $\ell_{\mathbf{c}}$ que es la tangente a $\mathcal{C}$ en $\mathbf{c}$; es decir, la línea de $\mathbf{a}$ a $\mathbf{c}$ es tangente al círculo.

Nótese que el argumento anterior es mucho más general: demuestra que para cualesquiera dos puntos $\mathbf{a}$ y $\mathbf{b}$ (distintos de $\mathbf{p}$) se tiene que

$$\mathbf{a} \in \ell_{\mathbf{b}} \Leftrightarrow \mathbf{b} \in \ell_{\mathbf{a}} . \tag{2.7}$$

Y los puntos del círculo son los únicos para los cuales se cumple $\mathbf{a} \in \ell_{\mathbf{a}}$.

En particular, hemos aprendido a calcular los puntos de tangencia a un círculo desde un punto exterior $\mathbf{a}$. A saber, de la ecuación lineal de su polar, $\ell_{\mathbf{a}}$, se despeja alguna de las dos variables y se sustituye en la ecuación del círculo. Esto nos da una ecuación de segundo grado en la otra variable que se puede resolver, y nos da dos raíces. Sustituyéndolas de nuevo en la ecuación de la polar se obtiene el otro par de coordenadas.

Por ejemplo, la ecuación del círculo (2.3) tiene la expresión vectorial

$$((x,y)-(3,-1))\cdot((x,y)-(3,-1))=4.$$

Si queremos conocer los puntos de tangencia desde el punto $\mathbf{a}=(1,3)$, encontramos primero la ecuación de su línea polar:

$$\begin{aligned} ((x,y)-(3,-1))\cdot((1,3)-(3,-1)) &= 4 \\ (x-3,y+1)\cdot(-2,4) &= 4 \\ -2x+4y+10 &= 4 \\ x-2y &= 3. \end{aligned}$$

De aquí, para encontrar $\ell_{\mathbf{a}}\cap\mathcal{C}$, conviene sustituir $x=3+2y$ en la ecuación original del círculo (2.3) para obtener

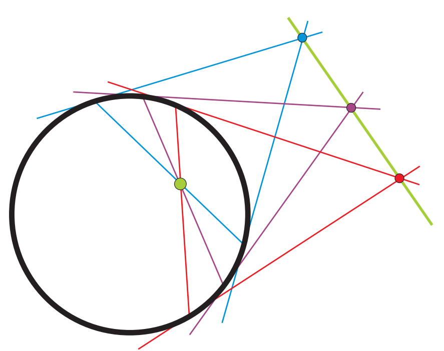
$$\begin{aligned} (3+2y)^2+y^2-6(3+2y)+2y &= -6 \\ 5y^2+2y-3 &= 0. \end{aligned}$$

Las raíces de esta ecuación cuadrática se obtienen por la fórmula clásica que los estudiantes de la ciudad de México llaman "del chicharronero":[1]

$$y=\frac{-2\pm\sqrt{4+60}}{10}=\frac{-2\pm 8}{10},$$

que nos da los valores $y_0=-1$ y $y_1=3/5$. Y estos, al sustituir de nuevo en la ecuación de la polar nos dan los puntos de tangencia de $\mathbf{a}$; que son $(1,-1)$ y $(1/5)(21,3)$. El estudiante puede verificar que satisfacen la ecuación del círculo, aunque los hayamos obtenido (al final) sustituyendo en la ecuación lineal de la polar, y que efectivamente sus tangentes pasan por $\mathbf{a}$.

EJERCICIO 2.9 Encuentra los puntos de tangencia:
a) al círculo $x^2-2x+y^2-4y=-3$ desde el punto $(-1,2)$;
b) al círculo $x^2+y^2=1$ desde el punto $(2,2)$.



EJERCICIO 2.10 Sea σ una secante de un círculo $\mathcal{C}$, es decir un segmento con extremos en el círculo. Si trazamos las tangentes por sus extremos, éstas se intersectan en un punto, que llamaremos el *polar* de σ. Demuestra que tres secantes son concurrentes si y sólo si sus puntos polares son colineales.

[1] El polinomio ax^2+bx+c tiene como raíces a $(-b\pm\sqrt[2]{b^2-4ac})/2a$.

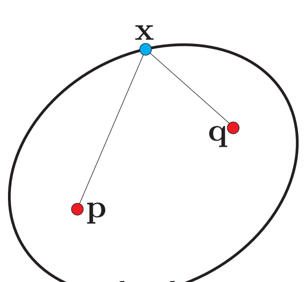
EJERCICIO 2.11 El estudiante que en el ejercicio anterior observó la necesidad formal de eliminar los diámetros en la definición de *punto polar* de un segmento, hizo muy bien; pero el que no se puso quisquilloso, y ni cuenta se dio, hizo mejor. Pensó proyectivamente. Demuestra que si permitimos que el centro $\mathbf{p}$ entre a nuestras consideraciones anteriores y añadimos puntos "polares" a los diámetros que juntos forman la nueva línea $\ell_{\mathbf{p}}$ polar con $\mathbf{p}$, entonces la polaridad se extiende a todas las líneas (falta definirla en las tangentes y exteriores) y que es una biyección entre líneas y puntos que cambia concurrencia por colinearidad. Observa además que los nuevos puntos (polares de los diámetros) tienen una dirección (no orientada) bien definida, aquella en la que se alejan los puntos polares de una secante que se acerca al diámetro correspondiente.

EJERCICIO 2.12 Demuestra que si $\mathbf{c}$ es un punto exterior (al círculo $\mathcal{C}$ con centro $\mathbf{p}$) entonces su recta a $\mathbf{p}$ bisecta sus dos tangentes a $\mathcal{C}$. Y además que las distancias a sus *pies* en $\mathcal{C}$ (es decir, a los puntos de tangencia) son iguales.

EJERCICIO 2.13 Observa que toda nuestra discusión sobre líneas polares se basó en la ecuación vectorial (2.4) que tiene sentido en $\mathbb{R}^3$ donde define esferas. Define la noción de plano polar a un punto respecto de una esfera en $\mathbb{R}^3$. Demuestra que un punto $\mathbf{c}$ en el exterior tiene como plano polar a uno que intersecta a la esfera en un círculo formado por los puntos de la esfera cuyas líneas a $\mathbf{c}$ son tangentes (a la unión de estas líneas se le llama el *cono* de la esfera con ápice $\mathbf{c}$).

2.2 Elipses

Si amarramos un burro hambriento a una estaca en un prado, se irá comiendo el pasto deambulando al azar, pero terminará por dejar pelón el interior de un círculo (con centro la estaca y radio la longitud del mecate). Si amarramos los extremos del mecate a dos estacas y le engarzamos al burro, lo que dibujará es una elipse; por eso, a fijar dos tachuelas en un papel, amarrar holgadamente un hilo entre ellas y luego tensar con un lápiz y girarlo, se le conoce como "*el método del jardinero*" para trazar elipses. Formalmente, las *elipses* son el lugar geométrico de los puntos (posiciones de la boca del burro cuando tensa el mecate) cuya suma de distancias a dos puntos fijos llamados *focos* (las estacas) es constante (la longitud del mecate). De tal manera que una elipse $\mathcal{E}$ queda determinada por la ecuación



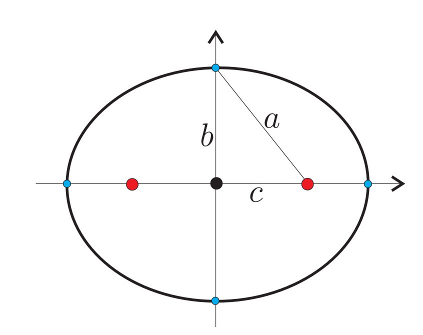
$$d(\mathbf{x},\mathbf{p}) + d(\mathbf{x},\mathbf{q}) = 2a, \tag{2.8}$$

donde $\mathbf{p}$ y $\mathbf{q}$ son los focos y a es una constante positiva, llamada *semieje mayor*, tal que $2a > d(\mathbf{p},\mathbf{q})$.

Incluimos el coeficiente 2 en la constante para que quede claro que si los focos coinciden, $\mathbf{p} = \mathbf{q}$; entonces se obtiene un círculo de radio a y centro en el foco; así que los círculos son elipses muy especiales. Esta ecuación, poniéndole coordenadas a

los focos, incluye raíces cuadradas por las distancias, y es aún muy "fea". Veremos ahora que en un caso especial es equivalente a una ecuación cuadrática "bonita".

Supongamos que el *centro* de la elipse $\mathcal{E}$, *i.e.*, el punto medio entre los focos, está en el origen y que además los focos están en el eje x. Entonces tenemos que $\mathbf{p} = (c, 0)$ y $\mathbf{q} = (-c, 0)$ para alguna c tal que $0 < c < a$ (donde ya suponemos que la elipse no es un círculo al pedir $0 < c$). Es fácil ver entonces que la intersección de $\mathcal{E}$ con el eje x consiste en los puntos $(a, 0)$ y $(-a, 0)$, pues la ecuación (2.8) para puntos $(x, 0)$ es



$$|x - c| + |x + c| = 2a,$$

que sólo tiene las soluciones $x = a$ y $x = -a$ (véase el Ejercicio 2.15), y de aquí el nombre de *semieje mayor* para la constante a. Como el eje y es ahora la mediatriz de los focos, en él, es decir en los puntos $(0, y)$, la ecuación (2.8) se vuelve

$$\sqrt{c^2 + y^2} = a,$$

que tiene soluciones $y = \pm b$, donde $b > 0$, llamado el *semieje menor* de la elipse $\mathcal{E}$, es tal que

$$b^2 = a^2 - c^2.$$

Ahora sí, consideremos la ecuación (2.8), que con $\mathbf{x} = (x, y)$ y la definición de nuestros focos, se expresa

$$\sqrt{(x - c)^2 + y^2} + \sqrt{(x + c)^2 + y^2} = 2a. \tag{2.9}$$

Si elevamos al cuadrado directamente esta ecuación, en el lado izquierdo nos quedaría un incómodo término con raíces. Así que conviene pasar una de las dos raíces al otro lado, para obtener

$$\sqrt{(x - c)^2 + y^2} = 2a - \sqrt{(x + c)^2 + y^2}.$$

Al elevar al cuadrado, se tiene

$$\begin{aligned} (x - c)^2 + y^2 &= 4a^2 - 4a\sqrt{(x + c)^2 + y^2} + (x + c)^2 + y^2 \qquad (2.10)\\ x^2 - 2cx + c^2 &= 4a^2 - 4a\sqrt{(x + c)^2 + y^2} + x^2 + 2cx + c^2 \\ 4a\sqrt{(x + c)^2 + y^2} &= 4a^2 + 4cx \\ a\sqrt{(x + c)^2 + y^2} &= a^2 + cx \end{aligned}$$

Elevando de nuevo al cuadrado, nos deshacemos de la raíz, y después, al agrupar términos, obtenemos

$$\begin{aligned} a^2\left((x+c)^2+y^2\right) &= a^4+2a^2cx+c^2x^2 \qquad (2.11)\\ a^2x^2+2a^2cx+a^2c^2+a^2y^2 &= a^4+2a^2cx+c^2x^2\\ \left(a^2-c^2\right)x^2+a^2y^2 &= a^2\left(a^2-c^2\right)\\ b^2x^2+a^2y^2 &= a^2b^2, \end{aligned}$$

que, al dividir entre a^2b^2, se escribe finalmente como

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1, \qquad (2.12)$$

llamada la *ecuación canónica de la elipse.*

Hemos demostrado que si $\mathbf{x}=(x,y)$ satisface (2.8), entonces satisface (2.12). Lo inverso, aunque también es cierto, no es tan automático, y requiere más argumentación pues en dos ocaciones (al pasar a las ecuaciones (2.10) y (2.11)) elevamos al cuadrado. Es cierto que si dos números son iguales entonces también sus cuadrados lo son (que fue lo que usamos); pero si $r^2=t^2$ no necesariamente se tiene que $r=t$, pues podría ser que $r=-t$. Éste es el problema conceptual conocido como "la raíz cuadrada tiene dos ramas", entonces los pasos de las ecuaciones (2.10) y (2.11) hacia arriba requieren argumentación sobre los signos (todos los demás pasos hacia arriba sí son automáticos). A manera de verificación, demostremos el inverso directamente.

Supongamos que $\mathbf{x}=(x,y)$ satisface la ecuación canónica de la elipse (2.12); veremos que entonces satisface la ecuación (2.9). Como en (2.12) se tiene una suma de números positivos que dan 1, entonces ambos sumandos están entre 0 y 1. Por lo tanto $|x|\leq|a|=a$, y esto habrá que usarse al sacar una raíz. De la ecuación (2.12) se puede despejar y^2 y sustituirla en las expresiones dentro de las raíces de (2.9), para obtener, en la primera,

$$\begin{aligned} (x-c)^2+y^2 &= (x-c)^2+b^2-\frac{b^2}{a^2}x^2\\ &= \left(1-\frac{b^2}{a^2}\right)x^2-2cx+c^2+b^2\\ &= \left(\frac{a^2-b^2}{a^2}\right)x^2-2cx+\left(c^2+b^2\right)\\ &= \frac{c^2}{a^2}x^2-2cx+a^2=\left(\frac{c}{a}x-a\right)^2. \end{aligned}$$

Ahora bien, como $|x|\leq|a|=a$ implica que $x\leq a$, y además $0<c<a$, entonces $cx<a^2$ y por tanto $\frac{c}{a}x<a$. Así que la expresión que pusimos arriba es negativa y entonces la raíz es la otra posible rama; es decir, hemos demostrado que

$$\sqrt{(x-c)^2+y^2}=a-\frac{c}{a}x.$$

De manera análoga, pero usando ahora que $-x \leq a$, se demuestra que

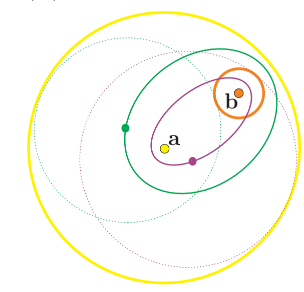
$$\sqrt{(x+c)^2 + y^2} = a + \frac{c}{a}x,$$

de donde se sigue la ecuación de las distancias (2.9).

EJERCICIO 2.14 Dibuja algunas elipses usando el método del jardinero. ¿Verdad que no salen de un solo trazo?

EJERCICIO 2.15 Sea $c > 0$. Dibuja la gráfica de la función $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = |x - c| + |x + c|$. Concluye que si $a > c$ entonces $f(x) = 2a$ si y sólo si $|x| = a$.

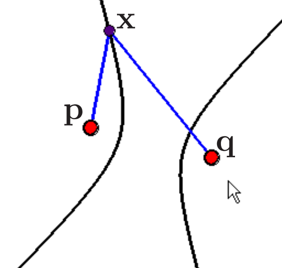
EJERCICIO 2.16 Sean $\mathcal{C}_1$ y $\mathcal{C}_2$ dos círculos con centros $\mathbf{a}$ y $\mathbf{b}$ respectivamente, tales que $\mathcal{C}_2$ está contenido en el interior de $\mathcal{C}_1$. Demuestra que el conjunto de puntos que son centro de un círculo $\mathcal{C}_3$ tangente tanto a $\mathcal{C}_1$ como a $\mathcal{C}_2$ es un par de elipses con focos $\mathbf{a}$ y $\mathbf{b}$. Observa que en el caso límite en que $\mathcal{C}_1$ es tangente a $\mathcal{C}_2$, una de las elipses se degenera en un segmento.



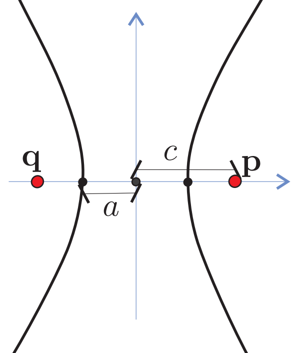
2.3 Hipérbolas

La *hipérbola* está definida como el lugar geométrico de los puntos cuya diferencia (en valor absoluto) de sus distancias a dos puntos fijos $\mathbf{p}$ y $\mathbf{q}$, llamados focos, es constante. Entonces, una hipérbola $\mathcal{H}$ está definida por la ecuación

$$|d(\mathbf{x}, \mathbf{p}) - d(\mathbf{x}, \mathbf{q})| = 2a, \tag{2.13}$$



donde $a > 0$, y además $2a < d(\mathbf{p}, \mathbf{q}) =: 2c$ (véase el Ejercicio 2.17). Si tomamos como focos a $\mathbf{p} = (c, 0)$ y $\mathbf{q} = (-c, 0)$, esta ecuación toma la forma



$$\left|\sqrt{(x-c)^2 + y^2} - \sqrt{(x+c)^2 + y^2}\right| = 2a. \tag{2.14}$$

Veremos a continuación que es equivalente a

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \tag{2.15}$$

donde $b > 0$ está definida por $a^2 + b^2 = c^2$. A esta última ecuación se le llama la *ecuación canónica de la hipérbola*.

Como las dos primeras ecuaciones involucran valores absolutos, entonces tienen dos posibilidades que corresponden a las dos ramas de la hipérbola. En una de ellas la distancia a uno de los focos es mayor y en la otra se invierten los papeles. De la ecuación (2.14) se tienen dos posibilidades:

$$\begin{aligned}\sqrt{(x-c)^2+y^2} &= 2a+\sqrt{(x+c)^2+y^2}\\ \sqrt{(x+c)^2+y^2} &= 2a+\sqrt{(x-c)^2+y^2};\end{aligned} \tag{2.16}$$

la primera corresponde a la rama donde $x<0$ y la segunda a $x>0$. Como se hizo en el caso de la elipse (elevando al cuadrado, simplificando para aislar la raíz que queda y luego volviendo a elevar al cuadrado y simplificando, véase Ejercicio 2.18) se obtiene, de cualquiera de las dos ecuaciones anteriores, la ecuación

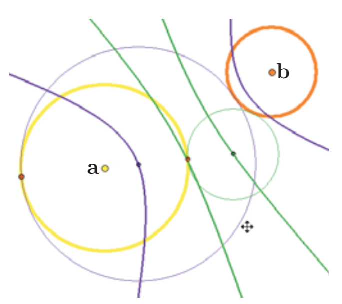
$$\left(a^2-c^2\right)x^2+a^2y^2=a^2\left(a^2-c^2\right). \tag{2.17}$$

Y de aquí, al sustituir $-b^2=a^2-c^2$, y dividir entre $-a^2b^2$ se obtiene la ecuación canónica.

EJERCICIO 2.17 Demuestra que la ecuación (2.13) define la mediatriz de $\mathbf{p}$ y $\mathbf{q}$ para $a=0$; los rayos complementarios del segmento $\overline{\mathbf{pq}}$ para $a=c$, y el conjunto vacío para $a>c$.

EJERCICIO 2.18 Demuestra que la ecuación (2.17) se sigue de cualquiera de las dos anteriores (2.16).

EJERCICIO 2.19 Demuestra que si (x,y) satisface la ecuación (2.15) entonces $|x|\geq a$. Concluye (sustituyendo en las raíces, como en el caso de la elipse) que, entonces, alguna de las dos ecuaciones (2.16) se satisfacen.



EJERCICIO 2.20 Sean $\mathcal{C}_1$ y $\mathcal{C}_2$ dos círculos con centros $\mathbf{a}$ y $\mathbf{b}$ respectivamente, y sea $\mathcal{X}$ el conjunto de los puntos que son centro de un círculo $\mathcal{C}_3$ tangente tanto a $\mathcal{C}_1$ como a $\mathcal{C}_2$.

a) Demuestra que si $\mathcal{C}_1$ y $\mathcal{C}_2$ están fuera uno del otro y sus radios son distintos, entonces $\mathcal{X}$ consta de un par de hipérbolas con focos $\mathbf{a}$ y $\mathbf{b}$.

b) Demuestra que si $\mathcal{C}_1$ y $\mathcal{C}_2$ se intersectan pero no son tangentes y sus radios son distintos, entonces $\mathcal{X}$ consta de un hipérbola y una elipse con focos $\mathbf{a}$ y $\mathbf{b}$.

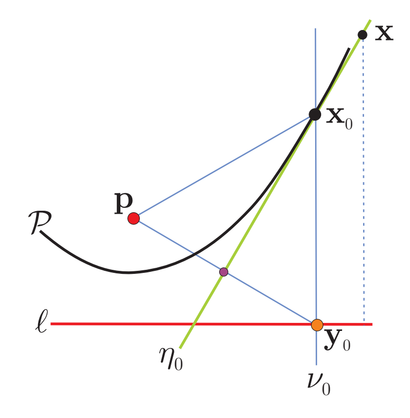
c) Discute los casos límite entre los casos anteriores y el del Ejercicio 2.16 y el de radios iguales.

2.4 Parábolas

Una *parábola* es el lugar geométrico de los puntos que equidistan de un punto $\mathbf{p}$ (llamado su *foco*) y una recta ℓ, llamada su *directriz*, donde $\mathbf{p}\notin\ell$. Es decir, está

definida por la ecuación
$$d(\mathbf{x},\mathbf{p}) = d(\mathbf{x},\ell) .$$
Tomemos un ejemplo sencillo con el foco en el eje y, la directriz paralela al eje x, y que además pase por el origen. Tenemos entonces $\mathbf{p} = (0, c)$, donde $c > 0$ digamos, y $\ell : y = -c$; de tal manera que la parábola queda determinada por la ecuación
$$\sqrt{x^2 + (y-c)^2} = |y + c| .$$
Como ambos lados de la ecuación son positivos, ésta es equivalente a la igualdad de sus cuadrados que da
$$\begin{aligned} x^2 + y^2 - 2cy + c^2 &= y^2 + 2cy + c^2 \\ x^2 &= 4cy. \end{aligned}$$
De tal manera que la gráfica de la función x^2 ($y = x^2$) es una parábola con foco $(0, 1/4)$ y directriz $y = -1/4$. Veremos ahora que cualquier parábola cumple "la propiedad de ser gráfica de una función" respecto de su directriz.

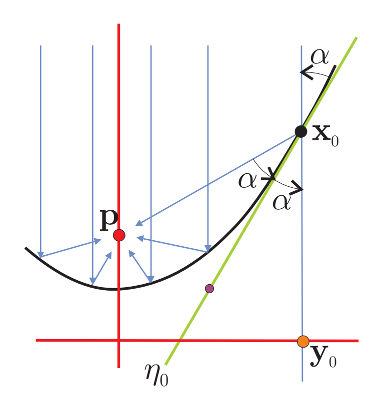
Sea $\mathcal{P}$ la parábola con foco $\mathbf{p}$ y directriz ℓ (donde $\mathbf{p} \notin \ell$). Dado un punto $\mathbf{y}_0 \in \ell$, es claro que los puntos del plano cuya distancia a ℓ coincide con (o se mide por) su distancia a $\mathbf{y}_0$ son precisamente los de la normal a ℓ que pasa por $\mathbf{y}_0$, llamémosla ν_0. Por otro lado, la mediatriz entre $\mathbf{y}_0$ y $\mathbf{p}$, llamémosla η_0, consta de los puntos cuyas distancias a $\mathbf{p}$ y a $\mathbf{y}_0$ coinciden. Por lo tanto la intersección de η_0 y ν_0 está en la parábola $\mathcal{P}$, es decir, $\mathbf{x}_0 = \nu_0 \cap \eta_0 \in \mathcal{P}$, pues $d(\mathbf{x}_0, \ell) = d(\mathbf{x}_0, \mathbf{y}_0) = d(\mathbf{x}_0, \mathbf{p})$. Pero además $\mathbf{x}_0$ es el único punto en la normal ν_0 que está en $\mathcal{P}$. Ésta es "la propiedad de la gráfica" a la que nos referíamos.



Podemos concluir aún más: que la mediatriz η_0 es la tangente a $\mathcal{P}$ en $\mathbf{x}_0$. Pues para cualquier otro punto $\mathbf{x} \in \eta_0$ se tiene que su distancia a ℓ es menor que su distancia a $\mathbf{y}_0$ que es su distancia a $\mathbf{p}$ ($d(\mathbf{x},\ell) < d(\mathbf{x},\mathbf{y}_0) = d(\mathbf{x},\mathbf{p})$), y entonces $\mathbf{x} \notin \mathcal{P}$. De hecho, la parábola $\mathcal{P}$ parte el plano en dos pedazos, los puntos más cerca de $\mathbf{p}$ que de ℓ (lo de adentro, digamos, definido por la desigualdad $d(\mathbf{x},\mathbf{p}) \leq d(\mathbf{x},\ell)$) y los que están más cerca de ℓ que de $\mathbf{p}$ (lo de afuera, dado por $d(\mathbf{x},\ell) \leq d(\mathbf{x},\mathbf{p})$ en donde está η_0), que comparten la frontera donde estas distancias coinciden (la parábola $\mathcal{P}$). Así que η_0 pasa *tangente* a $\mathcal{P}$ en $\mathbf{x}_0$, pues $\mathbf{x}_0 \in \eta_0 \cap \mathcal{P}$ y además η_0 se queda de un lado de $\mathcal{P}$.

2.5 Propiedades focales

2.5.1 De la parábola



Del análisis geométrico anterior es fácil deducir la *propiedad focal* de la parábola que la ha hecho tan importante en la tecnología de los siglos recientes, pero que se conoce desde que Apolonio de Pergas (c. 255-170 A.C.) escribió "*Sobre las cónicas*".

Como la mediatriz η_0 es bisectriz del ángulo $\mathbf{p}\mathbf{x}_0\mathbf{y}_0$, que llamaremos 2α, entonces un fotón (partícula de luz) que viaja por ν_0 (la normal a ℓ) hacia ℓ y "dentro" de la parábola incide en ella con ángulo α (este ángulo se mide infinitesimalmente, es decir, con la tangente η_0); como el ángulo de incidencia es igual al de reflexión, entonces rebota en el rayo que va directo al foco $\mathbf{p}$.

Llamemos *eje* de la parábola a la perpendicular a la directriz que pasa por el foco. Se tiene entonces que un haz de fotones que viaja en rayos paralelos al eje se va reflejando en la parábola para convertirse en un haz que converge en el foco $\mathbf{p}$.

Una antena parabólica casera (que se obtiene girando en su eje a una parábola) apuntada a un satélite, refleja entonces la onda de radio que viene de éste (el satélite está tan lejos que esta onda ya "es" paralela) y la hace converger en el receptor (puesto en el foco): "medio metro" cuadrado de la onda se concentra simultáneo en el receptor puntual; tal amplificación hace fácil captarla. El proceso inverso se usa en lámparas o en antenas transmisoras, y reflejan un haz que emana del foco en un haz paralelo al eje.

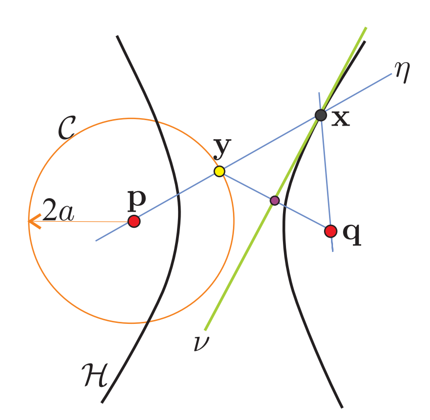
2.5.2 De la hipérbola

¿Qué pasa si plateamos (hacemos espejo) el otro lado de la parábola? Es fácil ver que entonces el haz paralelo al eje que incide en ella (por fuera) se refleja en un haz que emana (o simula emanar) del foco. Si pensamos que un haz paralelo viene de un punto al infinito y queremos reproducir este fenómeno para un haz puntual (que emana de un punto) hay que usar la hipérbola.

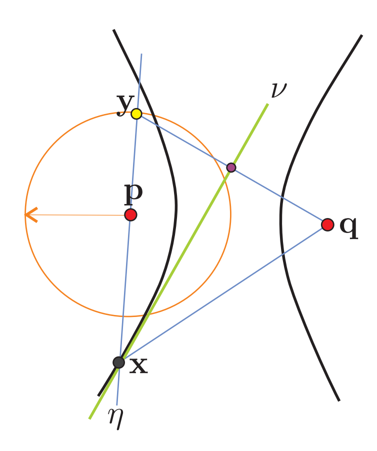
Sea $\mathcal{H}$ una hipérbola con focos $\mathbf{p}$ y $\mathbf{q}$; el haz que emana del foco $\mathbf{p}$ se refleja en la rama de $\mathcal{H}$ cercana a $\mathbf{q}$, en un haz que emana de $\mathbf{q}$. Entonces el uso tecnológico de la hipérbola es con refracción: si el fotón cruza la hipérbola con ángulo de incidencia igual al de refracción, entonces el haz que emana de $\mathbf{p}$ se refracta en uno confluyente en $\mathbf{q}$ (algo así se usa en el diseño de las lentes). Demostrar esto equivale a demostrar que la tangente a la hipérbola $\mathcal{H}$ en un punto $\mathbf{x}$ es bisectriz de sus líneas a los focos $\mathbf{p}$ y $\mathbf{q}$. Para demostrarlo usaremos una construcción análoga a la de la parábola pero cambiando el uso de las normales a la directriz (que "vienen" del infinito) por el haz de rayos que sale de $\mathbf{p}$.

Supongamos que $\mathcal{H}$ está acabada de determinar por la constante $2a$, es decir, que

$$\mathcal{H}: \quad |d(\mathbf{x}, \mathbf{p}) - d(\mathbf{x}, \mathbf{q})| = 2a,$$

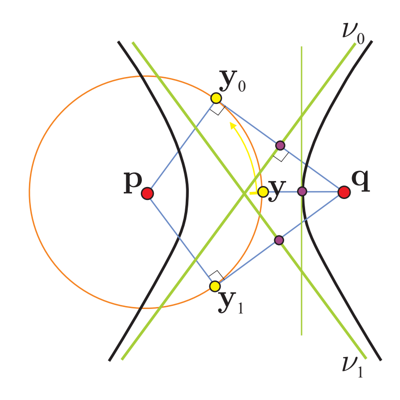


donde $0 < 2a < d(\mathbf{q}, \mathbf{p})$. Sea $\mathcal{C}$ el círculo con centro en $\mathbf{p}$ y radio $2a$, de tal manera que para cada $\mathbf{y} \in \mathcal{C}$ tenemos un rayo que sale de $\mathbf{p}$. Para fijar ideas, tomemos una $\mathbf{y} \in \mathcal{C}$ que apunta cerca de $\mathbf{q}$ (pero no justo a $\mathbf{q}$). Los puntos en el rayo $\vec{\eta}_{\mathbf{y}}$ que sale de $\mathbf{p}$ hacia $\mathbf{y}$ usan este rayo para medir su distancia a $\mathbf{p}$. Si tomamos a $\nu_{\mathbf{y}}$ como la mediatriz de $\mathbf{y}$ y el otro foco $\mathbf{q}$, entonces (como en la parábola) se tiene que $\mathbf{x} := \vec{\eta}_{\mathbf{y}} \cap \nu_{\mathbf{y}} \in \mathcal{H}$ pues $d(\mathbf{x}, \mathbf{q}) = d(\mathbf{x}, \mathbf{y})$ (porque $\mathbf{x} \in \nu_{\mathbf{y}}$) y $d(\mathbf{x}, \mathbf{p}) = d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{p}) = d(\mathbf{x}, \mathbf{y}) + 2a$ (donde estamos suponiendo que $\mathbf{x}$ está en el rayo que sale de $\mathbf{p}$ y pasa por $\mathbf{y}$ como en la figura), de tal manera que de las dos condiciones obtenemos que $d(\mathbf{x}, \mathbf{p}) - d(\mathbf{x}, \mathbf{q}) = 2a$.

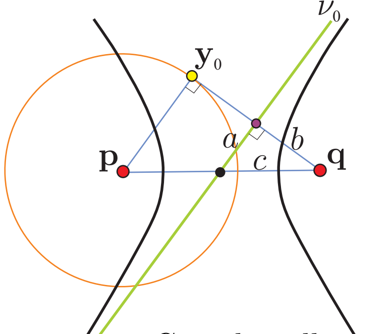


Antes de seguir adelante, consideremos el otro caso, a saber, cuando el punto de intersección $\mathbf{x}$ de las rectas $\eta_{\mathbf{y}}$ (por $\mathbf{p}$ y $\mathbf{y} \in \mathcal{C}$) y $\nu_{\mathbf{y}}$, la mediatriz de $\mathbf{y}$ y $\mathbf{q}$, está del otro lado de $\mathbf{y}$ (es decir, cuando $\mathbf{p}$ está en el segmento $\overline{\mathbf{xy}}$). Se tiene entonces que $d(\mathbf{x}, \mathbf{q}) = d(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}, \mathbf{p}) + d(\mathbf{p}, \mathbf{y}) = d(\mathbf{x}, \mathbf{p}) + 2a$, de tal manera que $\mathbf{x}$ está en la otra rama de la hipérbola.

Hemos demostrado que si $\mathbf{y} \in \mathcal{C}$ es tal que su recta por él y $\mathbf{p}$, $\eta_{\mathbf{y}}$ y su mediatriz con $\mathbf{q}$, $\nu_{\mathbf{y}}$, se intersectan en un punto $\mathbf{x}$, entonces $\mathbf{x} \in \mathcal{H}$. Como en el caso de la parábola, la mediatriz $\nu_{\mathbf{y}}$ es ahora el candidato a ser la tangente a $\mathcal{H}$ en el punto $\mathbf{x}$. Esto se demuestra viendo que $\left|d\left(\mathbf{x}', \mathbf{p}\right) - d\left(\mathbf{x}', \mathbf{q}\right)\right| < 2a$ para cualquier otro punto $\mathbf{x}' \in \nu_{\mathbf{y}}$ y se deja como ejercicio (hay que usar dos veces la desigualdad del triángulo). De aquí también se sigue la propiedad focal de la hipérbola.



Pero además, se sigue otro punto clave. La existencia de las asíntotas. Hay un par de puntos en el círculo $\mathcal{C}$, para los cuales la mediatriz con $\mathbf{q}$ es paralela a su recta por $\mathbf{p}$. Para que esto suceda para un punto $\mathbf{y} \in \mathcal{C}$, como la mediatriz con $\mathbf{q}$, $\nu_{\mathbf{y}}$, es ortogonal a la recta $\mathbf{qy}$, entonces es necesario (y suficiente) que el radio $\mathbf{py}$ sea también ortogonal a $\mathbf{qy}$; es decir, que $\mathbf{qy}$ sea la tangente al círculo $\mathcal{C}$ en el punto $\mathbf{y}$. Entonces las dos tangentes al círculo $\mathcal{C}$ que se pueden trazar desde $\mathbf{q}$, definen dos puntos, $\mathbf{y}_0$ y $\mathbf{y}_1$ digamos, cuyas mediatrices con $\mathbf{q}$, ν_0 y ν_1, son las *asíntotas* de $\mathcal{H}$.



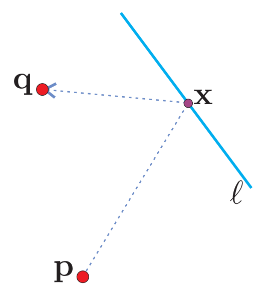
Para fijar ideas consideremos a los focos $\mathbf{p}$ y $\mathbf{q}$ en una línea horizontal y con $\mathbf{p}$ a la izquierda, como en las figuras. Si movemos el punto $\mathbf{y} \in \mathcal{C}$, empezando con el más cercano a $\mathbf{q}$, hacia arriba (contra el reloj), el correspondiente punto $\mathbf{x} \in \mathcal{H}$ viaja por la rama de $\mathbf{q}$ alejándose arriba a la derecha. Su tangente (que empieza vertical) es mediatriz de un segmento que gira en $\mathbf{q}$ siguiendo a su otro extremo $\mathbf{y}$; entonces la tangente gira hacia la derecha, como el reloj.

Cuando $\mathbf{y}$ llega a $\mathbf{y}_0$, el punto $\mathbf{x}$ (que viaja en $\mathcal{H}$) se ha "ido al infinito" pero su tangente llega plácidamente a ν_0, la asíntota correspondiente. Al mover otro poco a $\mathbf{y}$, la tangente $\nu_{\mathbf{y}}$ empieza a girar contra el reloj (pues el segmento $\mathbf{qy}$ empieza a bajar) y el correspondiente punto $\mathbf{x} \in \mathcal{H}$ reaparece pero en la otra rama y justo por el lado opuesto. Podemos seguir girando a $\mathbf{y}$ sin problemas, mientras $\mathbf{x}$ recorre toda la rama de $\mathbf{p}$, hasta llegar al otro punto de tangencia $\mathbf{y}_1$ (ahora abajo) y tenemos que $\mathbf{x}$ desaparece (en el infinito arriba), justo cuando la tangente está en la otra asíntota. Al seguir girando, $\mathbf{x}$ vuelve a cambiar de rama y la tangente de dirección de giro para regresar, junto con $\mathbf{y}$, a su lugar de origen.

Además, obtenemos la interpretación geométrica de los tres parámetros a, b, c en el triángulo rectángulo que se forma con el centro, el eje que une los dos focos y una de las asíntotas.

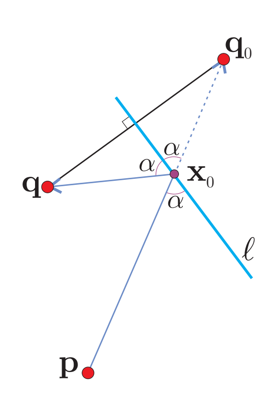
2.5.3 De la elipse

La propiedad focal de la elipse es que cualquier fotón que emana de un foco se refleja en la elipse para llegar al otro foco. De tal manera que si susurramos en uno de los focos de un cuarto elíptico, el sonido rebota en la pared y confluye estruendosamente en el otro foco (pues, además, cada onda sonora que emana de un foco llega simultáneamente al otro foco, sin distorsión o *delay*, pues en todas sus trayectorias ha recorrido la misma distancia). Una manera de ver esto es análoga a la que usamos para la hipérbola: tomar el círculo de radio $2a$ centrado en el foco $\mathbf{p}$ —este círculo contiene el otro foco $\mathbf{q}$, puesto que ahora $2a > d(\mathbf{p}, \mathbf{q})$—, y luego ver que para los puntos de este círculo, su mediatriz con $\mathbf{q}$ es tangente a la elipse $\mathcal{E}$. Pero para variarle, se la dejamos al estudiante y usamos otro método basado en el clásico "problema del bombero".

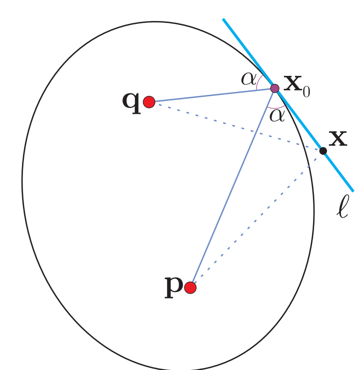


Supongamos que un bombero está parado en el punto $\mathbf{p}$ y hay un fuego —que tiene que apagar— en el punto $\mathbf{q}$. Pero tiene su cubeta vacía, y entonces tiene que pasar primero a llenarla a un río cuyo borde es la recta ℓ. El problema es ¿cuál es la trayectoria óptima que debe seguir el bombero? Es decir, para cuál punto $\mathbf{x} \in \ell$ se tiene que $d(\mathbf{p}, \mathbf{x}) + d(\mathbf{x}, \mathbf{q})$ es mínima. En la vida real, y con muchos bomberos en fila, se aproximarían a la solución poco a poco. Pero desde nuestra cómoda banca de la abstracción, hay una solución muy elegante.

No hemos especificado de qué lado del río está el fuego. Si estuviera del otro lado que el bombero, cualquier trayectoria al fuego tiene que pasar por ℓ y entonces debe irse por la línea recta de $\mathbf{p}$ a $\mathbf{q}$ y tomar agua en $\mathbf{x}_0 = \ell \cap \overline{\mathbf{pq}}$. Entonces, si fuego y bombero están del mismo lado del río ℓ, podemos pensar en un "fuego virtual", que es el reflejado de $\mathbf{q}$ en ℓ, llamémosle $\mathbf{q}_0$, que cumple que $d(\mathbf{x}, \mathbf{q}) = d(\mathbf{x}, \mathbf{q}_0)$ para todo $\mathbf{x} \in \ell$, (para $\mathbf{q}$ y $\mathbf{q}_0$, ℓ es su mediatriz). La solución es, por el caso anterior, $\mathbf{x}_0 = \ell \cap \overline{\mathbf{pq}_0}$. Pero obsérvese además que el ángulo α con el que llega el bombero a ℓ es igual al ángulo "de reflexión" con el que sale corriendo al fuego (ya con la cubeta llena), e igual al ángulo con el que seguiría su trayecto al fuego virtual, y que está propiedad determina al punto de mínimo recorrido $\mathbf{x}_0$; es fácil convencerse de que para cualquier otro punto de ℓ los ángulos de llegada y de salida son distintos. Si los bomberos fueran fotones que emanan de $\mathbf{p}$ y ℓ es un espejo, el único que llega a $\mathbf{q}$ es el fotón que sigue el recorrido mínimo.

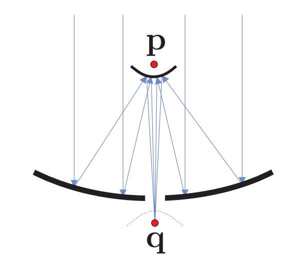


Pensemos ahora que $\mathbf{p}$ y $\mathbf{q}$ son los focos de una elipse y $\mathbf{x}_0$ un punto en ella. Sea ℓ la recta que pasa por $\mathbf{x}_0$ y bisecta (por fuera) los segmentos de $\mathbf{p}$ y $\mathbf{q}$ a $\mathbf{x}_0$. Por construcción y la solución al problema del bombero, cualquier otro punto $\mathbf{x} \in \ell$ tiene mayor suma de distancias a los focos y por tanto está fuera de la elipse. Esto demuestra que ℓ es la tangente a la elipse en el punto $\mathbf{x}_0$ y, por tanto, su propiedad focal con la que empezamos esta sección.



2.5.4 Telescopios

Por último, veamos brevemente cómo funciona un telescopio moderno. El haz de fotones que viene de una galaxia, digamos, es "paralelo" pues está tan lejos que desde la Tierra es imposible medir variación de ángulos (mucho más que los satélites respecto a la superficie de las antenas parbólicas). Un gran espejo parabólico hace confluir este haz en su foco $\mathbf{p}$. Pero $\mathbf{p}$ está del otro lado del espejo. Esto se resuelve poniendo un pequeño espejo secundario elíptico y con el mismo foco $\mathbf{p}$, que vuelve a reflejar el haz. Éste pasa por un pequeño agujero en el centro del primario y vuelve a confluir en su otro foco $\mathbf{q}$, que ahora sí está detras del gran espejo primario, y ahí se captura en el detector. El hoyo que se hace al primario, así como lo que tapa el secundario o la estructura, cuenta poco pues la superficie crece como el radio al cuadrado.



Otra manera de resolver el problema de "regresar el haz" hacia atrás del espejo primario, que es la más usada, es poniendo un secundario hiperbólico en vez de uno elíptico, con la superficie reflejante por fuera y otra vez con el mismo foco $\mathbf{p}$. De nuevo hace reflejar el haz en uno que confluye en su otro foco $\mathbf{q}$, cómodamente situado detras

del hoyito del primario.

Lo que cuenta es qué tantos fotones podemos capturar en el sistema óptico; a simple vista es la superficie de la pupila, los telescopios modernos recolectan los fotones que chocarían en decenas de metros cuadrados en una dirección dada y los hacen pasar, despues de dos reflexiones adecuadas, por la pupila de algún instrumento detector.

2.6 *Armonía y excentricidad

Hemos definido elipses e hipérbolas como lugares geométricos con base en las propiedades aditivas de distancias. Ahora estudiaremos las cónicas respecto a propiedades multiplicativas. Pues resulta que las tres cónicas (elipses, parábolas e hipérbolas) se definen también como el lugar geométrico de los puntos x cuya razón de sus distancias a un foco p y a una recta ℓ, llamada *directriz*, es una constante fija e, llamada su *excentricidad*. Es decir, están dadas por la ecuación

$$d(x, p) = e\, d(x, \ell).$$

Para $e = 1$, ésta es la ecuación de la parábola que ya estudiamos. Veremos que para $e < 1$ se define una elipse y para $e > 1$ una hipérbola, de tal manera que las tres cónicas quedan en una sola familia. Pero notemos que no apareció el círculo. Éste tiene excentricidad 0 que no cabe en la definición anterior (más que como límite). Sin embargo, aparece en una instancia (en apariencia) más simple de esta ecuación que consiste en tomar otro punto en vez de la línea, y veremos que ligado a esto surge naturalmente el concepto clásico de los puntos armónicos.

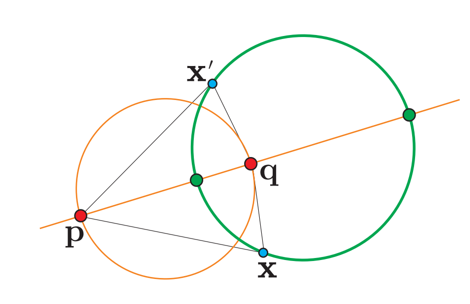
2.6.1 *Puntos armónicos y círculos de Apolonio

Sean p y q dos puntos distintos en el plano. Consideremos la ecuación

$$d(x, p) = h\, d(x, q), \tag{2.18}$$

donde $h > 0$. Para $h = 1$ ésta es la ecuación de la mediatriz que estudiamos en el capítulo anterior, pero a excepción de $h = 1$ donde da una recta, veremos que esta ecuación define círculos: los llamados *"círculos de Apolonio"*.[2]

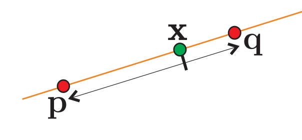
[2] También a un círculo tangente a tres dados se le llama *"círculo de Apolonio"*.



Notemos primero que el lugar geométrico $\mathcal{C}$ definido por (2.18) es simétrico respecto a la línea ℓ que pasa por $\mathbf{p}$ y $\mathbf{q}$, pues si un punto $\mathbf{x}$ satisface la ecuación también lo hace su reflejado $\mathbf{x}'$ respecto a la recta ℓ. De tal manera que la recta ℓ debe ser un diámetro y el círculo debe estar determinado por los puntos de intersección $\ell \cap \mathcal{C}$. Encontrémoslos.

Sabemos que los puntos de ℓ se expresan en coordenadas baricéntricas como $\mathbf{x} = \alpha\mathbf{p} + \beta\mathbf{q}$ con $\alpha + \beta = 1$, donde, si damos una dirección a ℓ, de $\mathbf{p}$ hacia $\mathbf{q}$ digamos, y denotamos a las distancias dirigidas con $\vec{d}$, se tiene que

$$\alpha = \frac{\vec{d}\,(\mathbf{x},\mathbf{q})}{\vec{d}\,(\mathbf{p},\mathbf{q})} \quad \text{y} \quad \beta = \frac{\vec{d}\,(\mathbf{x},\mathbf{p})}{\vec{d}\,(\mathbf{q},\mathbf{p})} = \frac{\vec{d}\,(\mathbf{p},\mathbf{x})}{\vec{d}\,(\mathbf{p},\mathbf{q})}.$$



De tal manera que si $\mathbf{x} = \alpha\mathbf{p} + \beta\mathbf{q} \in \ell$, entonces

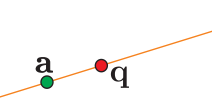
$$\frac{\beta}{\alpha} = \frac{\vec{d}\,(\mathbf{p},\mathbf{x})}{\vec{d}\,(\mathbf{x},\mathbf{q})}.$$

Por tanto, para encontrar las soluciones de (2.18) en ℓ basta encontrar las soluciones de $|\beta/\alpha| = h$ con $\alpha+\beta = 1$: al sustituir $\beta = h\alpha$ en esta última ecuación, encontramos una solución dentro del segmento $\overline{\mathbf{pq}}$ (con α y β positivas) dada por

$$\alpha_h := \frac{1}{h+1} \quad \text{y} \quad \beta_h := \frac{h}{h+1}.$$

Llamemos $\mathbf{a}$ al punto correspondiente, es decir, sea $\mathbf{a} = \alpha_h\mathbf{p} + \beta_h\mathbf{q}$. Obsérvese que $\mathbf{a}$ es el punto medio cuando $h = 1$; que $\mathbf{a}$ está más cerca de $\mathbf{p}$ para $h < 1$ y que se aproxima a $\mathbf{q}$ conforme h crece. Para $h \neq 1$, también tenemos otra solución fuera del segmento (que viene de tomar $\beta = -h\alpha$), a saber:

$$\alpha'_h = \frac{1}{1-h} = \frac{\alpha_h}{\alpha_h - \beta_h} \quad \text{y} \quad \beta'_h = \frac{h}{h-1} = \frac{\beta_h}{\beta_h - \alpha_h}.$$

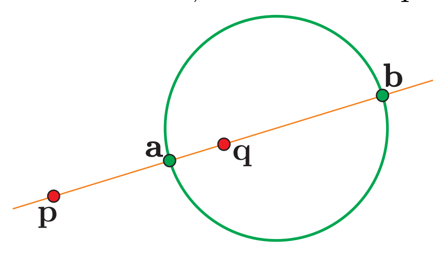


Llamemos $\mathbf{b}$ a esta otra solución de (2.18) en ℓ (es decir, $\mathbf{b} = \alpha'_h\mathbf{p} + \beta'_h\mathbf{q}$), y nótese que $\mathbf{b}$ está en el rayo exterior de ℓ correspondiente al punto, $\mathbf{p}$ o $\mathbf{q}$, más cercano a $\mathbf{a}$ ($h < 1$ o $h > 1$, respectivamente). (Y nótese también que la expresión de α_h y β_h en términos de α'_h y β'_h es exactamente la misma.)

Clásicamente, se dice que la pareja $\mathbf{a}, \mathbf{b}$ es *armónica* respecto a la pareja $\mathbf{p}, \mathbf{q}$ (o que $\mathbf{a}$ y $\mathbf{b}$ son *conjugados armónicos* respecto a $\mathbf{p}$ y $\mathbf{q}$) pues (como reza la definición general) los cuatro puntos están alineados y cumplen la relación

$$\frac{d\,(\mathbf{p},\mathbf{a})}{d\,(\mathbf{a},\mathbf{q})} = \frac{d\,(\mathbf{p},\mathbf{b})}{d\,(\mathbf{b},\mathbf{q})}, \tag{2.19}$$

que, en nuestro caso, es la *"razón de armonía"* h. Se tiene claramente que la relación de armonía es simétrica (pues $\mathbf{p}, \mathbf{q}$ son armónicos respecto a $\mathbf{a}, \mathbf{b}$). Y para aclarar otro poco su relación geométrica, hay que observar que si en la recta real consideramos a $\mathbf{p}$ y $\mathbf{q}$ como los puntos -1 y 1, entonces el conjugado armónico de x es x^{-1}, y el 0 queda como conjugado armónico del infinito (así como el punto medio de $\mathbf{p}$ y $\mathbf{q}$, $h = 1$, fue un caso especial).



Ahora sí, demostremos que el círculo cuyo diámetro es el segmento $\overline{\mathbf{ab}}$ está dado por la ecuación (2.18), que es, recuérdese, $d(\mathbf{x}, \mathbf{p}) = h\, d(\mathbf{x}, \mathbf{q})$. En el Ejercicio 2.7 se debió demostrar que el círculo con diametro $\overline{\mathbf{ab}}$ está dado por la ecuación $(\mathbf{x} - \mathbf{a}) \cdot (\mathbf{x} - \mathbf{b}) = 0$ (si no se hizo, compruébese ahora). Desarrollemos entonces esta expresión, basándonos en que los coeficientes son coordenadas baricéntricas (*i.e.*, que suman 1):

$$\begin{aligned}
(\mathbf{x} - \mathbf{a}) \cdot (\mathbf{x} - \mathbf{b}) &= (\mathbf{x} - \alpha_h \mathbf{p} - \beta_h \mathbf{q}) \cdot (\mathbf{x} - \alpha'_h \mathbf{p} - \beta'_h \mathbf{q}) \\
&= (\alpha_h (\mathbf{x} - \mathbf{p}) + \beta_h (\mathbf{x} - \mathbf{q})) \cdot (\alpha'_h (\mathbf{x} - \mathbf{p}) + \beta'_h (\mathbf{x} - \mathbf{q})) \\
&= \alpha_h \alpha'_h (\mathbf{x} - \mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}) + \beta_h \beta'_h (\mathbf{x} - \mathbf{q}) \cdot (\mathbf{x} - \mathbf{q}) \\
&\quad + (\alpha_h \beta'_h + \beta_h \alpha'_h) (\mathbf{x} - \mathbf{p}) \cdot (\mathbf{x} - \mathbf{q}) \qquad (2.20) \\
&= \left(\frac{1}{1 - h^2} \right) \left((\mathbf{x} - \mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}) - h^2 (\mathbf{x} - \mathbf{q}) \cdot (\mathbf{x} - \mathbf{q}) \right),
\end{aligned}$$

donde hemos usado las definiciones —en términos de $h \neq 1$— de nuestros coeficientes (véase el Ejercicio 2.21). Se tiene entonces que $(\mathbf{x} - \mathbf{a}) \cdot (\mathbf{x} - \mathbf{b}) = 0$ si y sólo si $(\mathbf{x} - \mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}) = h^2 (\mathbf{x} - \mathbf{q}) \cdot (\mathbf{x} - \mathbf{q})$, que es justo el cuadrado de la ecuación (2.18).

Además, tenemos al menos otras tres maneras interesantes de detectar armonía.

Teorema 2.1 *Sean $\mathbf{p}$, $\mathbf{q}$, $\mathbf{a}$, $\mathbf{b}$ cuatro puntos colineales, y sean $\mathcal{C}_1$ y $\mathcal{C}_2$ los círculos con diametro $\overline{\mathbf{pq}}$ y $\overline{\mathbf{ab}}$ respectivamente. Entonces son equivalentes:*

i) *Las parejas $\mathbf{p}, \mathbf{q}$ y $\mathbf{a}, \mathbf{b}$ son armónicas.*

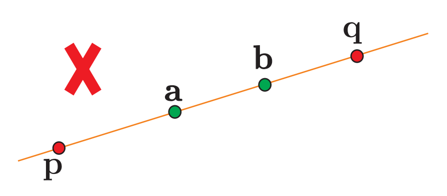
ii) *Se cumple que $(\mathbf{a} - \mathbf{p}) \cdot (\mathbf{b} - \mathbf{q}) = -(\mathbf{a} - \mathbf{q}) \cdot (\mathbf{b} - \mathbf{p})$.*

iii) *El punto $\mathbf{b}$ está en la línea polar de $\mathbf{a}$ respecto a $\mathcal{C}_1$.*

iv) *Los círculos $\mathcal{C}_1$ y $\mathcal{C}_2$ se intersectan ortogonalmente.*

Demostración. **(i)⇔(ii).** La condición de armonía (2.19) claramente se puede reescribir, usando normas para expresar distancias, como

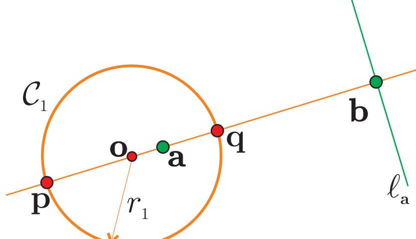
$$|\mathbf{a} - \mathbf{p}|\, |\mathbf{b} - \mathbf{q}| = |\mathbf{a} - \mathbf{q}|\, |\mathbf{b} - \mathbf{p}|. \qquad \text{(i)}$$



La equivalencia de (i) y **(ii)** se sigue del hecho de que si dos vectores $\mathbf{u}$ y $\mathbf{v}$ son paralelos entonces $|\mathbf{u}|\,|\mathbf{v}| = \pm\,(\mathbf{u}\cdot\mathbf{v})$, dependiendo de si apuntan en la misma dirección $(+)$ o en direcciones opuestas $(-)$. Esto da inmediatamente la implicación **(ii)**⇒(i), pues por la hipótesis general del teorema los cuatro puntos están alineados y entonces sus diferencias son paralelas. Para el regreso sólo hay que tener cuidado con los signos. Si se cumple (i) es fácil ver que el orden lineal en que aparecen los puntos en la línea es alternado (entre las parejas), pues si, por ejemplo, $\mathbf{p}$ y $\mathbf{q}$ fueran los extremos, un lado de (i) es estrictamente menor que el otro. Entonces —renombrando entre las parejas si es necesario— podemos suponer que aparecen en el orden $\mathbf{p}$, $\mathbf{a}$, $\mathbf{q}$, $\mathbf{b}$. En este caso (i) toma la expresión **(ii)**.

(i)⇒(iii). Supongamos que $\mathbf{p}$, $\mathbf{q}$ y $\mathbf{a}$, $\mathbf{b}$ son los de los párrafos anteriores con razón de armonía $h \neq 1$. Sean $\mathbf{o}$ (ojo, no necesariamente es el origen 0) el centro y r_1 el radio del círculo $\mathcal{C}_1$, es decir, $\mathbf{o} = (1/2)(\mathbf{p}+\mathbf{q})$ y $r_1 = |\mathbf{p}-\mathbf{q}|/2$. De tal manera que $\mathcal{C}_1 : \ (\mathbf{x}-\mathbf{o})\cdot(\mathbf{x}-\mathbf{o}) = r_1^2$.

Por la sección 1.1, ver que $\mathbf{b}$ está en la línea polar de $\mathbf{a}$ respecto a $\mathcal{C}_1$ equivale a demostrar que $(\mathbf{a}-\mathbf{o})\cdot(\mathbf{b}-\mathbf{o}) = r_1^2$. Pero ya evaluamos $(\mathbf{x}-\mathbf{a})\cdot(\mathbf{x}-\mathbf{b}) = (\mathbf{a}-\mathbf{x})\cdot(\mathbf{b}-\mathbf{x})$ para cualquier $\mathbf{x}$; entonces hay que sustituir $\mathbf{o}$ en (2.20), para obtener



$$
\begin{aligned}
(\mathbf{a}-\mathbf{o})\cdot(\mathbf{b}-\mathbf{o}) &= \left(\frac{1}{1-h^2}\right)\big((\mathbf{o}-\mathbf{p})\cdot(\mathbf{o}-\mathbf{p}) \\
&\quad -h^2\,(\mathbf{o}-\mathbf{q})\cdot(\mathbf{o}-\mathbf{q})\big) \\
&= \left(\frac{1}{1-h^2}\right)\left(r_1^2 - h^2 r_1^2\right) = r_1^2.
\end{aligned}
$$

Por el papel que desempeñó (y desempeñará) el punto $\mathbf{o}$, conviene pensar que sí es el origen. Si no lo fuera, lo trasladamos al origen junto con el resto de los puntos y las propiedades que nos interesan se preservan. Puede pensarse también que en lo que sigue cada letra ($\mathbf{b}$ digamos) representa al punto original menos $\mathbf{o}$ (a $\mathbf{b}-\mathbf{o}$ en el ejemplo). Tenemos entonces que $\mathbf{p} = -\mathbf{q}$ (pues el origen es ahora su punto medio) y que $r_1^2 = \mathbf{p}\cdot\mathbf{p}$. Ahora la condición **(iii)** equivale a

$$\mathbf{a}\cdot\mathbf{b} = r_1^2, \tag{iii}$$

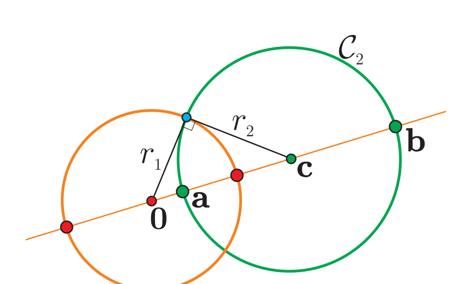
y podemos demostrar la equivalencia de **(ii)** y **(iii)** de un solo golpe (aunque un lado ya esté demostrado).

(ii)⇔(iii). Usando que $\mathbf{q} = -\mathbf{p}$, claramente se tiene

$$(\mathbf{a}-\mathbf{p})\cdot(\mathbf{b}-\mathbf{q}) + (\mathbf{a}-\mathbf{q})\cdot(\mathbf{b}-\mathbf{p}) = 2\,(\mathbf{a}\cdot\mathbf{b} - \mathbf{p}\cdot\mathbf{p}),$$

que concluye la demostración (**(ii)** es que el lado izquierdo sea cero y (iii) que el lado derecho lo es).

(iii)⇔(iv). Sean $\mathbf{c} = (1/2)(\mathbf{a} + \mathbf{b})$ el centro de $\mathcal{C}_2$, y $r_2 = |\mathbf{b} - \mathbf{a}|/2$ su radio. Que $\mathcal{C}_1$ y $\mathcal{C}_2$ sean *ortogonales* quiere decir que sus tangentes en el punto de intersección lo son; y esto equivale a que sus radios al punto de intersección son ortogonales (pues las tangentes son ortogonales a los radios). Por el Teorema de Pitágoras esto equivale a que



$$r_1^2 + r_2^2 = |\mathbf{c}|^2 \qquad \text{(iv)}$$

(pues $\mathcal{C}_1$ está centrado en el origen, así que $|\mathbf{c}|$ es la distancia entre los centros), que es lo que tomamos en adelante como (**iv**). Sabemos, por hipótesis, que $\mathbf{a}$ y $\mathbf{b}$ son paralelos y hay que observar que ambas condiciones ($\mathbf{a} \cdot \mathbf{b} = r_1^2 > 0$ y $r_1^2 + r_2^2 = |\mathbf{c}|^2$) implican que están del mismo lado del origen. Podemos suponer entonces que $\mathbf{a}$ es "más pequeño" que $\mathbf{b}$, y entonces

$$\begin{aligned} |\mathbf{c}|^2 &= (|\mathbf{a}| + r_2)^2 \\ &= |\mathbf{a}|(|\mathbf{a}| + 2r_2) + r_2^2 \\ &= |\mathbf{a}||\mathbf{b}| + r_2^2 \\ &= \mathbf{a} \cdot \mathbf{b} + r_2^2. \end{aligned}$$

Lo cual demuestra la equivalencia (iii)⇔(iv) y concluye el Teorema. □

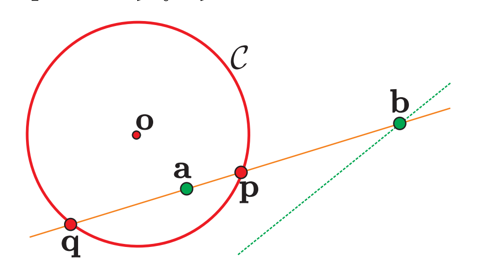
En resumen, los círculos de Apolonio de un par de puntos $\mathbf{p}$ y $\mathbf{q}$ son los ortogonales al círculo que los tiene como diámetro y cuyo centro está en esa línea, y además la cortan en sus conjugados armónicos; teniendo como límites los "círculos de radio cero" en $\mathbf{p}$ y en $\mathbf{q}$ y su línea mediatriz.

EJERCICIO 2.21 Demuestra que $\alpha_h \beta'_h = -\beta_h \alpha'_h$ (usando la definición de los correspondientes primados sin referencia a h); y usando las definiciones con base en h, demuestra que

$$\alpha_h \alpha'_h = \frac{1}{1 - h^2} \quad \text{y} \quad \beta_h \beta'_h = \frac{h^2}{h^2 - 1}$$

son las coordenadas baricéntricas del centro del círculo (2.18) respecto a $\mathbf{p}$ y $\mathbf{q}$.

EJERCICIO 2.22 Sean $\mathbf{p}$ y $\mathbf{q}$ dos puntos en un círculo $\mathcal{C}$ con centro $\mathbf{o}$; y sea $\mathbf{a}$ un punto en el segmento $\overline{\mathbf{pq}}$ distinto del punto medio, es decir, $\mathbf{a} = \alpha\mathbf{p} + \beta\mathbf{q}$ con $\alpha + \beta = 1$, $0 < \alpha < 1$ y $\alpha \neq \beta$. Demuestra que el punto armónico de $\mathbf{a}$ respecto a $\mathbf{p}$ y $\mathbf{q}$ (llamémoslo $\mathbf{b}$) está en la línea polar de $\mathbf{a}$ respecto al círculo $\mathcal{C}$. (Expresa $\mathbf{b}$ en coordenadas baricéntricas respecto a $\mathbf{p}$ y $\mathbf{q}$, y sustituye en la ecuación de la polar de $\mathbf{a}$). Concluye que la línea polar de un punto $\mathbf{a}$ en el interior de un círculo consiste de todos sus armónicos respecto a la intersección con el círculo de las líneas que pasan por el.



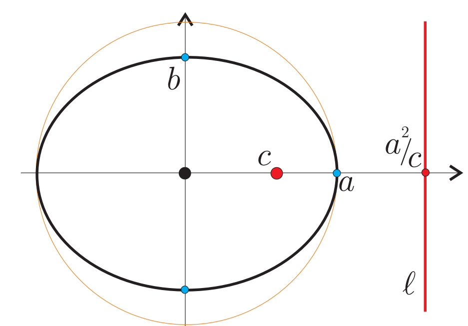
2.6.2 *Excentricidad

Demostraremos ahora que elipses e hipérbolas se definen por la ecuación

$$d(\mathbf{x}, \mathbf{p}) = e\, d(\mathbf{x}, \ell)\,, \tag{2.21}$$

con $e < 1$ y $e > 1$ respectivamente, donde $\mathbf{p}$ es uno de sus focos, y a ℓ la llamaremos su *directriz* correspondiente. La constante e se llama la *excentricidad* de la cónica, y ya sabemos que las de excentricidad 1 son las parábolas.

Consideremos la elipse canónica $\mathcal{E}$ con centro en el origen y semiejes $a > b$ en los ejes coordenados. Sabemos que sus focos están en $\mathbf{p} = (c, 0)$ y $\mathbf{q} = (-c, 0)$, donde $c > 0$ es tal que $b^2 + c^2 = a^2$. Sabemos además que $\mathcal{E}$ intersecta al eje x, que es su eje focal, en los puntos $(a, 0)$ y $(-a, 0)$.



Resulta que la directriz de $\mathcal{E}$ correspodiente al foco $\mathbf{p}$ es ortogonal al eje x y lo intersecta en el conjugado armónico de $\mathbf{p}$ respecto a los puntos de intersección; es decir, es la línea polar a $\mathbf{p}$ respecto al círculo $\mathbf{x}\cdot\mathbf{x} = a^2$ (que, notémoslo, es doblemente tangente a $\mathcal{E}$). Sea entonces ℓ definida por la ecuación

$$x = \frac{a^2}{c}.$$

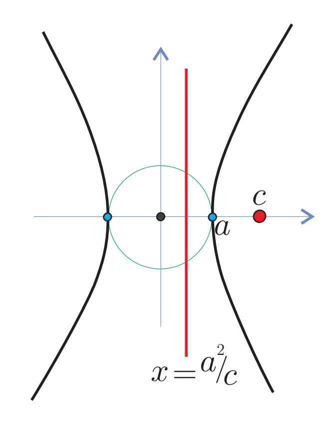
De los datos que tenemos podemos deducir su excentricidad (pues del punto $(a, 0) \in \mathcal{E}$ conocemos sus distancias a $\mathbf{p}$ y ℓ), que resulta $e = c/a < 1$. Desarrollemos entonces a la ecuación (2.21) en coordenadas para ver si son ciertas las afirmaciones que nos hemos sacado de la manga de la historia.

Puesto que ambos lados de (2.21) son positivos, esta ecuación es equivalente a su cuadrado que en coordenadas (y nuestra definición de $\mathbf{p}, \ell$ y e) toma la forma

$$\begin{aligned}
(x - c)^2 + y^2 &= \left(\frac{c}{a}\right)^2 \left(x - \frac{a^2}{c}\right)^2 \\
x^2 - 2cx + c^2 + y^2 &= \left(\frac{c^2}{a^2}\right)\left(x^2 - 2\frac{a^2}{c}x + \frac{a^4}{c^2}\right) \\
\left(\frac{a^2 - c^2}{a^2}\right) x^2 + y^2 &= a^2 - c^2.
\end{aligned}$$

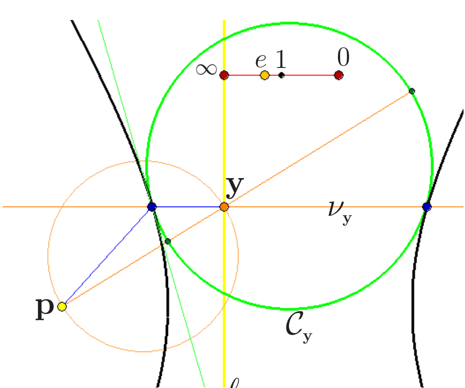
De aquí, al dividir entre $b^2 = a^2 - c^2$ se obtiene la ecuación canónica de la elípse $\mathcal{E}$ y quedan demostradas todas nuestras aseveraciones.

Para una hipérbola $\mathcal{H}$, de nuevo la directriz correspondiente a un foco es perpendicular al eje focal y lo intersecta en su conjugado armónico respecto a los puntos de intersección con $\mathcal{H}$. De tal manera que podemos tomar exactamente los mismos datos del caso anterior ($\mathbf{p} = (c, 0)$, $\ell : \ cx = a^2$ y $e = c/a$), y la ecuación (2.21) se desarrolla igual. Pero ahora tenemos que $c > a$ (la excentricidad es mayor que 1, $e = c/a > 1$) y entonces definimos $b > 0$ tal que $a^2 + b^2 = c^2$; de tal manera que en el último paso hay que dividir entre $-b^2$ para obtener



$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

La excentricidad de una cónica es un invariante de semejanza , es decir, si magnificamos (hacemos *zoom*) o disminuimos una cónica, su excentricidad se mantiene. Pues esta magnificación (o contracción) se logra alejando (o acercando) al foco $\mathbf{p}$ y a la directriz ℓ. Pero si fijamos a estos últimos y variamos la excentricidad e, se cubre por cónicas todo el plano (exceptuando a $\mathbf{p}$ y ℓ que aparecen como los límites $e \to 0$ y $e \to \infty$ respectivamente) pues cualquier punto da una razón específica entre las distancias y por tanto una excentricidad que lo incluye.



Recreemos la construcción de la parábola a partir de las mediatrices, pero ahora con círculos de Apolonio. Si $\mathcal{P}$ es la cónica con foco $\mathbf{p}$, directriz ℓ y excentricidad e (puede ser elipse, parábola o hipérbola), y para cada punto $\mathbf{y}$ en ℓ tomamos el círculo de Apolonio $\mathcal{C}_\mathbf{y}$ entre $\mathbf{p}$ y $\mathbf{y}$ con razón de armonía e, entonces está claro que $\mathcal{C}_\mathbf{y} \cap \nu_\mathbf{y} \subset \mathcal{P}$ (donde $\nu_\mathbf{y}$ es la normal a ℓ en $\mathbf{y}$). Pues la distancia de un punto en $\nu_\mathbf{y}$ a ℓ corresponde a su distancia a $\mathbf{y}$; y si está en $\mathcal{C}_\mathbf{y}$ la razón de las distancias a $\mathbf{p}$ y $\mathbf{y}$ es la adecuada. Pero ahora $\mathcal{C}_\mathbf{y} \cap \nu_\mathbf{y}$ no es necesariamente un punto pues, salvo el caso $e = 1$ donde $\mathcal{C}_\mathbf{y}$ es la mediatriz, un círculo intersecta a una recta en 2, 1 o 0 puntos.

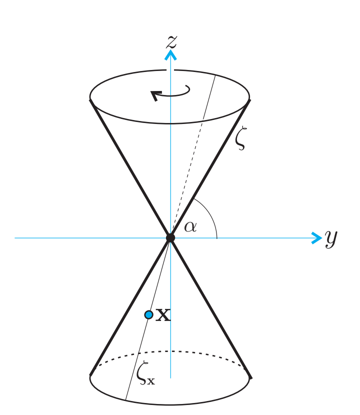
Para el caso $e < 1$, los círculos de Apolonio $\mathcal{C}_\mathbf{y}$ contienen a $\mathbf{p}$ en su interior; crecen conforme $\mathbf{y}$ se aleja en ℓ, pero llega un momento (que corresponde al semieje menor de la elipse) en el que $\nu_\mathbf{y}$ ya no los corta. Por el contrario, para $e > 1$ los círculos de Apolonio contienen a $\mathbf{y}$, y entonces $\nu_\mathbf{y}$ siempre los corta en dos puntos que dibujan las dos ramas de la hipérbola.

En ambos casos, si $\mathcal{C}_\mathbf{y}$ corta a $\nu_\mathbf{y}$ en $\mathbf{x}$, digamos, entonces por un argumento análogo al de la parábola se puede ver que está fuera de $\mathcal{P}$, es decir, que $\mathcal{C}_\mathbf{y}$ es *tangente* a $\mathcal{P}$ en $\mathbf{x}$, pero en este caso, y salvo por el último toque a una elipse en el semieje menor, también es tangente a $\mathcal{P}$ en otro punto. Para la hipérbola es fácil

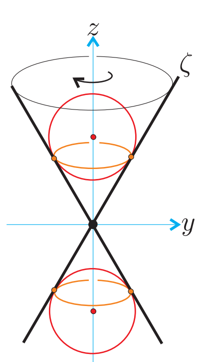
verlos, simplemente dejando caer un círculo hacia el centro y si es más grande que la cinturita: ahí donde se detiene es un $\mathcal{C}_y$, con y correspondiendo a la altura de los dos puntos de $\mathcal{P}$ donde se atoró.

2.7 *Esferas de Dandelin

Hemos desarrollado todo este capítulo a partir de las definiciones planas de las cónicas por propiedades de distancias. Sin embargo, en su introducción empezamos como los griegos: definiéndolas en el espacio como intersección de un cono con planos. Veremos brevemente que tal definición corresponde a la que usamos, basándonos en una demostración que data de cerca del año de 1800 debida a un tal Dandelin.

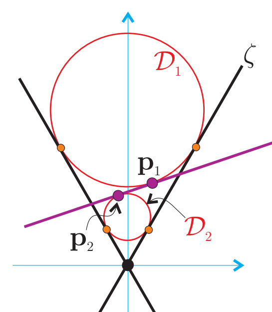


Para definir un cono $\mathcal{C}$ en $\mathbb{R}^3$, consideremos una línea ζ por el origen en el plano yz con un ángulo α respecto al eje y: y hagámosla girar alrededor del eje z. De tal manera que la intersección de $\mathcal{C}$ con un plano horizontal ($z = c$, digamos) es un círculo centrado en "el origen" ($(0, 0, c)$, y de radio $c \cos\alpha / \sin\alpha$). Por la definición, es claro que para cualquier punto $\mathbf{x} \in \mathcal{C}$, su recta por el origen, que denotaremos con $\zeta_\mathbf{x}$, está contenida en el cono, y la llamaremos *la recta del cono por* $\mathbf{x}$. Queremos hacer ver que cualquier plano Π que no contenga al origen intersecta el cono $\mathcal{C}$ en una cónica (definida por distancias).



Las *esferas tangentes* al cono son esferas con centro en el eje z y que tocan al cono pero que no lo cortan, es decir, que tienen a todas las líneas del cono como líneas tangentes. Si dejamos caer una esfera dentro del cono se asienta en una esfera tangente. Es claro que las esferas tangentes se obtienen al hacer girar en el eje z un círculo centrado en él y tangente a la línea generadora ζ; que para cada radio hay exactamente dos esferas tangentes (una arriba y una abajo), y que para cada punto del eje z hay una única esfera tangente centrada en él. Además, cada esfera tangente al cono lo intersecta en un círculo horizontal llamado su *círculo de tangencia* o de *contacto*.

Tomemos ahora un plano Π casi horizontal. Este plano Π tiene exactamente dos esferas tangentes que también son tangentes al cono $\mathcal{C}$, que llamaremos sus *esferas de Dandelin,* $\mathcal{D}_1$ y $\mathcal{D}_2$. Veamos y, para fijar ideas, pensemos que el plano Π intersecta el cono arriba del origen, pero no muy lejos. Si pensamos en una esfera muy grande tangente al cono, ésta está muy arriba; al desinflarla (manteniéndola siempre tangente al cono) va bajando. Hay un primer momento en el que toca a Π, ésa es $\mathcal{D}_1$. Al seguir desinflándola, intersecta a Π en círculos, y hay un último momento en que lo toca —esa es $\mathcal{D}_2$— para seguir decreciendo, insignificante, hacia el origen.



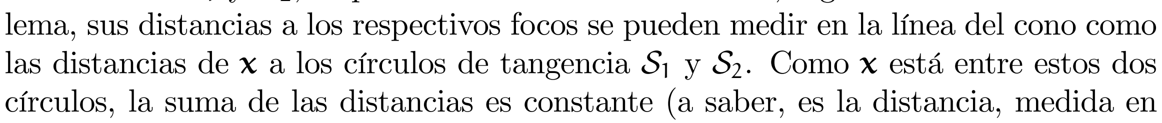
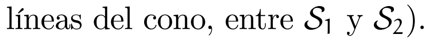
Más formalmente, y sin perder generalidad, podemos suponer que el plano Π contiene la dirección del eje x; entonces las esferas de Dandelin de Π corresponden a dos de los cuatro círculos tangentes a las tres rectas que se obtienen de intersectar el plano yz con el plano Π y el cono $\mathcal{C}$ (precisamente, los dos que caen dentro del cono).

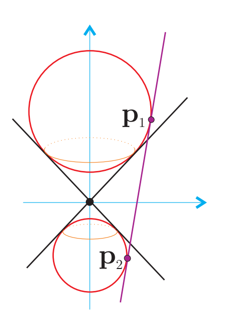
Sean $\mathbf{p}_1$ y $\mathbf{p}_2$ los puntos donde $\mathcal{D}_1$ y $\mathcal{D}_2$ tocan, respectivamente, el plano Π. Estos serán los focos de la elipse $\Pi \cap \mathcal{C}$. Para verlo necesitamos un lema que es más difícil enunciar que probar.

Lema 2.2 *Sea $\mathcal{D}$ una esfera de Dandelin de un plano Π. Sea $\mathbf{p}$ su punto de contacto con Π; y sea $\mathcal{S}$ el círculo de tangencia de $\mathcal{D}$ en $\mathcal{C}$. Para cualquier $\mathbf{x} \in \Pi \cap \mathcal{C}$, sea $\mathbf{y}$ la intersección de la recta del cono por $\mathbf{x}$ y el círculo de tangencia,* i.e. $\mathbf{y} = \mathcal{S} \cap \zeta_{\mathbf{x}}$*; se tiene entonces que*

$$d(\mathbf{x}, \mathbf{p}) = d(\mathbf{x}, \mathbf{y}).$$

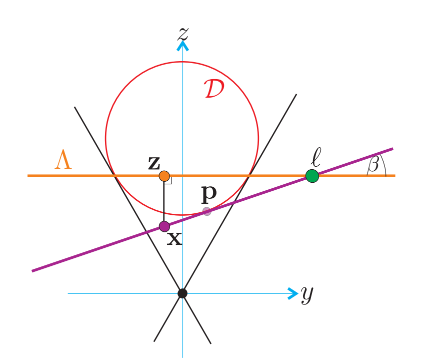
Basta observar que los segmentos $\overline{\mathbf{xp}}$ y $\overline{\mathbf{xy}}$ son ambos tangentes a la esfera $\mathcal{D}$ y que, en general, dos segmentos que salen del mismo punto y llegan tangentes a una esfera miden lo mismo. □

Ahora sí, podemos demostrar que $\Pi \cap \mathcal{C}$ es una elipse tal como la definimos. Sean $\mathcal{S}_1$ y $\mathcal{S}_2$ los círculos de tangencia de las esferas de Dandelin $\mathcal{D}_1$ y $\mathcal{D}_2$, respectivamente. Dado $\mathbf{x} \in \Pi \cap \mathcal{C}$, según el  lema, sus distancias a los respectivos focos se pueden medir en la línea del cono como las distancias de $\mathbf{x}$ a los círculos de tangencia $\mathcal{S}_1$ y $\mathcal{S}_2$. Como $\mathbf{x}$ está entre estos dos círculos, la suma de las distancias es constante (a saber, es la distancia, medida en  líneas del cono, entre $\mathcal{S}_1$ y $\mathcal{S}_2$).



Cuando el plano Π es cercano a un plano vertical (incluyendo este caso), se tiene que Π intersecta ambos lados del cono $\mathcal{C}$. Pero también tiene dos esferas de Dandelin tangentes a él y al cono. De nuevo, sus puntos de tangencia con el plano son los focos de la hipérbola $\Pi \cap \mathcal{C}$, y la demostración es análoga (usando el lema). Dejamos los detalles al lector (siguiente ejercicio), pues para el caso de la parábola necesitamos encontrar la directriz, y su análisis incluirá los tres casos.

Supongamos que el plano Π tiene un ángulo $\beta > 0$ respecto al plano horizontal xy, y que $\mathcal{D}$ es su esfera de Dandelin, es decir, que es tangente a Π y al cono $\mathcal{C}$ con punto y círculo de contacto $\mathbf{p}$ y $\mathcal{S}$ respectivamente. Sea Λ el plano horizontal que contiene al círculo de tangencia $\mathcal{S}$. Sea $\ell = \Pi \cap \Lambda$, línea que podemos pensar como paralela al eje x y que llamaremos la *directriz*.

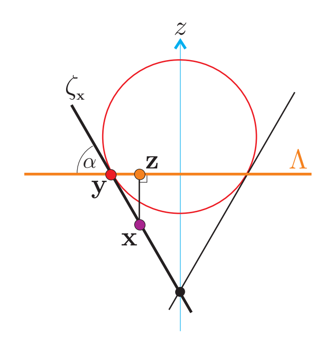


Consideremos un punto $\mathbf{x} \in \Pi \cap \mathcal{C}$. De la proyección al plano yz, es fácil deducir que

$$d(\mathbf{x}, \ell) \operatorname{sen} \beta = d(\mathbf{x}, \mathbf{z}),$$

donde $\mathbf{z}$ es la proyección ortogonal de $\mathbf{x}$ al plano horizontal Λ. Tomemos ahora el plano vertical que pasa por $\mathbf{x}$ y el eje z. Claramente contiene a $\mathbf{z}$ y a la línea del cono $\zeta_{\mathbf{x}}$ que pasa por $\mathbf{x}$; y es fácil ver en él que

$$d(\mathbf{x}, \mathbf{z}) = d(\mathbf{x}, \mathbf{y}) \operatorname{sen} \alpha,$$



donde $\mathbf{y} = \zeta_{\mathbf{x}} \cap \Lambda$ (justo la $\mathbf{y}$ que aparece en el lema). De estas dos ecuaciones y el lema se sigue que

$$d(\mathbf{x}, \mathbf{p}) = \frac{\operatorname{sen} \beta}{\operatorname{sen} \alpha} d(\mathbf{x}, \ell).$$

Esto demuestra que los planos que no pasan por el origen (los que tienen esferas de Dandelin) intersectan el cono $\mathcal{C}$ en cónicas cuya excentricidad depende de los ángulos (tal como queríamos demostrar).

Obsérvese que no eliminamos los planos horizontales ($\beta = 0$) de la afirmación anterior. Cuando el plano Π tiende a uno horizontal, la cónica se hace un círculo, el foco $\mathbf{p}$ tiende a su centro y la directriz ℓ se aleja hacia el infinito.

EJERCICIO 2.23 Concluye la demostración de que $\Pi \cap \mathcal{C}$ es una hipérbola cuando las dos esferas de Dandelin de Π están en los dos lados del cono usando a éstas y a los dos focos.

EJERCICIO 2.24 Demuestra que si $\beta = \alpha$, entonces el plano Π tiene una única esfera de Dandelin (el caso de la parábola).

Capítulo 3

Transformaciones

Sin duda, el concepto de función tiene un papel fundamental en todas las ramas de la matemática de hoy en día: en la geometría elemental (aunque sea antigua), se dice que dos figuras son *congruentes* si existe una *transformación rígida* que lleve una sobre la otra, y que son *semejantes* si existe una de estas transformaciones que seguida de una *homotecia* (un cambio de escala) lleve una a la otra; en el análisis o el cálculo se estudian las funciones *integrables* o *diferenciables*; en la topología, las funciones *continuas*; en la teoría de los grupos se estudian los *homomorfismos* (funciones entre grupos que preservan la operación ahí definida); en el álgebra lineal se estudian las funciones *lineales* y las *afines*, etc.; siempre que hay algún tipo interesante de "objetos matemáticos" parece haber una correspondiente noción de "funciones" que los relacionan.

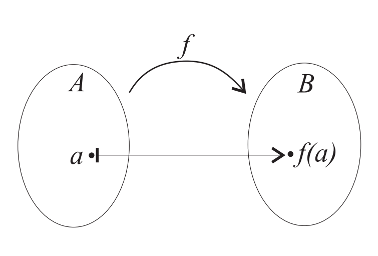
Sin haberlo hecho explícito, los griegos manejaban intuitivamente el concepto de transformación rígida y de semejanza; por ejemplo, en el axioma "todos los ángulos rectos son iguales", en la palabra "iguales" se incluye la idea de que se puede *mover* uno hasta traslaparse sobre el otro, y en sus teoremas de semejanza la noción ya se hace explícita y habla en el fondo de un cierto tipo de funciones. Más adelante, en el surgimiento del cálculo (Newton y Leibniz) así como en el de la geometría analítica (Descartes) o en el algebra de los arabes, las funciones tenían un papel importante pero debajo del agua o en casos muy concretos (funciones reales de variable real dadas por fórmulas en el cálculo, por ejemplo). Sin embargo, el aislamiento del concepto de función, la generalidad de la noción —que englobaba cosas al parecer distantes—, es muy reciente: termina de afinarse con la teoría de los conjuntos que arranca Cantor en la segunda mitad del siglo XIX. Pero es tan inmediata su aceptación (o bien, ya estaba tan madura su concepción) que a finales del siglo XIX, Felix Klein, en una famosísima conferencia (conocida como "El Programa de Erlangen"), se atreve a afirmar, basado en el trabajo de su gran amigo Sophus Lie, que la geometría es el estudio de un espacio (un conjunto de puntos, piénsese en el plano) junto con un *grupo de transformaciones* (un conjunto específico de funciones del espacio en sí mismo) y de las estructuras que permanecen *invariantes* bajo el grupo. En fin, todavía el estudiante no tiene ejemplos

claros de estas nociones y es casi imposible que lo aprecie. Pero el punto es que, de alguna manera, Klein dijo: para hacer geometría es importantísimo estudiar las *transformaciones* (funciones) del plano en sí mismo, conocerlas de arriba a abajo; y en este capítulo eso intentamos pues el siglo XX, al transcurrir, le fue dando más y más razón al visionario.

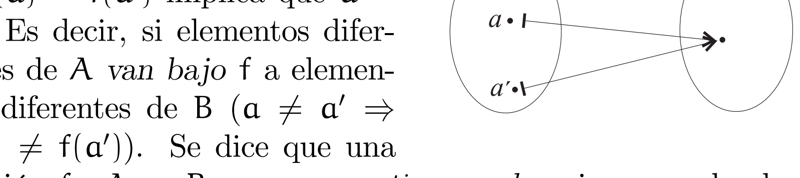
Damos muy rápidamente las nociones generales de función en la primera sección. Pues, aunque parece que se nos cae el nivel por un ratito, vale la pena establecer cierta terminología muy general que se usa en su manejo y ciertos ejemplos muy particulares que serán de gran interés en el estudio de las transformaciones geométricas y que, a su vez, servirán para familiarizarnos con los conceptos básicos. Después entramos a las transformaciones geométricas. El orden en que lo hacemos no es el que técnicamente facilita las cosas (empezar por transformaciones lineales), sino el que intuitivamente parece más natural (empezar por transformaciones rígidas); y de ahí, deducir la noción de transformación lineal para regresar de nuevo y desarrollar las fórmulas analíticas.

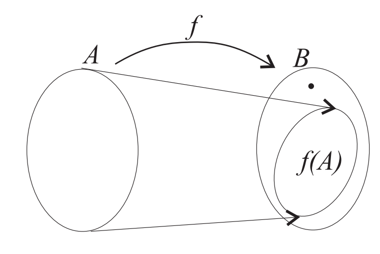
3.1 Funciones y transformaciones

En los siguientes párrafos se usa el tipo de letra *este* para las nociones que se definen formalmente y *este otro* para terminología cómoda y coloquial que facilita mucho el manejo de las funciones.



Dados dos conjuntos A y B, una *función* f de A a B, denotado $f : A \to B$, es una manera de *asociar* a cada elemento $a \in A$ un elemento de B, denotado $f(a)$. Por ejemplo, las funciones de $\mathbb{R}$ en $\mathbb{R}$ se describen comúnmente mediante fórmulas como $f(x) = x^2$ o $g(x) = 2x + 3$, que dan la regla para asociar a cada número otro número (por ejemplo, $f(2) = 4$ o $g(-1) = 1$). Otro ejemplo: en el capítulo 1 vimos las funciones de $\mathbb{R}$ en $\mathbb{R}^2$ que describen rectas parametrizadas. Pero en general, y ésa es la maravillosa idea generalizadora (valga el pleonasmo), una función no tiene por qué estar dada por una fórmula; a veces es algo dado por "Dios", una "caja negra", que "sabe" como asociar a los elementos del *dominio* (el conjunto A, en la notación con que empezamos) elementos del *contradominio* B.





Se dice que una función $f : A \to B$ es *inyectiva* si $f(a) = f(a')$ implica que $a = a'$. Es decir, si elementos diferentes de A *van bajo* f a elementos diferentes de B ($a \neq a' \Rightarrow f(a) \neq f(a')$). Se dice que una función $f : A \to B$ es *suprayectiva* o *sobre* si para cada elemento de B hay uno en A que le "*pega*", es decir, si para cada $b \in B$ existe $a \in A$ tal que $f(a) = b$. Y se dice que es *biyectiva* o una *biyección*,

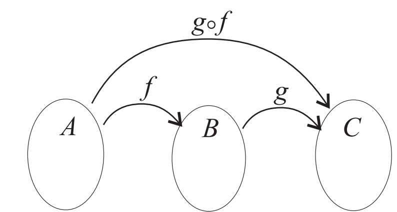
si es inyectiva y suprayectiva; también se le llama *correspondencia biunívoca.* (Diagramas de funciones *no* inyectiva y *no* sobre en las figuras adjuntas).

Las funciones tienen una noción natural de composición, al aplicarse una tras otra para dar una nueva función. Dadas dos funciones $f : A \to B$ y $g : B \to C$, su *composición*, denotada $g \circ f$ y a veces llamada f *seguida de* g o bien "g *bolita* f" es la función

$$g \circ f : A \to C$$

definida por la fórmula

$$(g \circ f)(a) = g(f(a)).$$



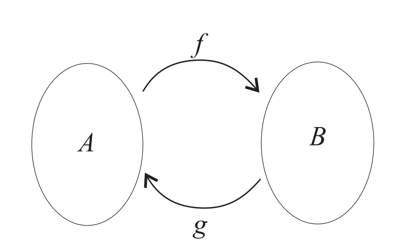
Hay que resaltar que la composición de funciones está definida sólo cuando el *dominio* de g (es decir, de donde "*sale*" o en donde está definida, el conjunto B en nuestro caso) es igual al *contradominio* de f (es decir, a donde "*llega*", el final de la flecha, donde "*caen*", otra vez B en nuestro caso); si no fuera así, "g no sabría que hacer con algunos $f(a)$". También hay que resaltar que la dirección de la escritura "g *bolita* f" es la contraria a la de la acción o lectura ("f seguida de g" o bien "f *compuesta con* g"); y esto se ha convenido así por la costumbre y puesto que la fórmula, que a final de cuentas es la que manda, queda mucho más natural.

Por ejemplo, con las funciones $f, g : \mathbb{R} \to \mathbb{R}$ que definimos con fórmulas unos párrafos antes se tiene que $(g \circ f)(x) = 2x^2 + 3$ mientras que $(f \circ g)(x) = 4x^2 + 12x + 9$, así que *sí* importa el orden; la composición está lejos de ser conmutativa; en general, aunque $g \circ f$ esté definida, $f \circ g$ ni siquiera tiene sentido.

Cada conjunto A, trae consigo una función llamada su *identidad* definida por

$$\begin{aligned} \mathrm{id}_A : A &\to A \\ \mathrm{id}_A(a) &= a, \end{aligned}$$

que es la función que no *hace* nada, que deja a todos en su lugar. Y, aunque parezca inocua, es fundamental darle un nombre, pues entonces podemos escribir mucho, por ejemplo:



Lema 3.1 *Dada una función* $f : A \to B$*, se tiene*

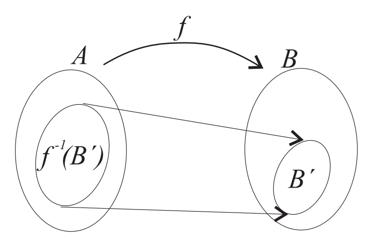
i) f *es inyectiva* $\Leftrightarrow$ $\exists$ $g : B \to A$ *tal que* $g \circ f = \mathrm{id}_A$,
ii) f *es suprayectiva* $\Leftrightarrow$ $\exists$ $g : B \to A$ *tal que* $f \circ g = \mathrm{id}_B$.

EJERCICIO 3.1 Da ejemplos de funciones inyectivas que no son sobre y a la inversa.

EJERCICIO 3.2 Demuestra el lema anterior y el siguiente corolario de él.

Corolario 3.2 *Dada una función* $f : A \to B$*, entonces* f *es biyectiva si y sólo si existe una función* $g : B \to A$ *tal que* $g \circ f = \mathrm{id}_A$ *y* $f \circ g = \mathrm{id}_B$*.*

En este caso, a g se le da el nombre de *inversa* de f y se le denota f^{-1}. Aunque aquí hay que hacer notar que el símbolo f^{-1} se usa también de una manera más general para denotar conjuntos. Pues para cualquier subconjunto $B' \subset B$ podemos definir su *imagen inversa o pre-imagen* $f^{-1}(B') \subset A$ como



$$f^{-1}(B') := \{a \in A \mid f(a) \in B'\} \ .$$

Tenemos entonces que f es inyectiva si y sólo si para todo $b \in B$ se tiene que $\sharp f^{-1}(b) \leq 1$ (donde $\sharp$ denota cardinalidad y hemos identificado $f^{-1}(b)$ con $f^{-1}(\{b\})$), es decir, "a cada elemento de B le pega a lo más uno de A"; y que f es sobre si y sólo si para todo $b \in B$ se tiene que $\sharp f^{-1}(b) \geq 1$ (es decir, que $f^{-1}(b)$ no es el *conjunto vacío*). Por lo tanto, f es biyectiva si y sólo si $\sharp f^{-1}(b) = 1$ para todo $b \in B$, es decir, si y sólo si f^{-1} es una función bien definida al aplicarla a *singuletes* de B.

También se puede definir la *imagen directa* de subconjuntos $A' \subset A$, o simplemente su *imagen*, como

$$f(A') := \{f(a) \mid a \in A'\} \subset B;$$

y la *imagen* de la función es $f(A)$, la imagen del dominio.

Finalmente llegamos a la definición que más nos interesa.

Definición 3.1.1 Una *transformación* de A es una función biyectiva de A en A.

Hay que hacer notar que el término "transformación" se usa de diferentes maneras en otros textos y en otros contextos. Pero aquí estaremos tan enfocados a funciones biyectivas de un conjunto en sí mismo que lo asignaremos a ellas.

EJERCICIO 3.3 Demuestra que si $f : A \to B$ y $g : B \to C$ son biyectivas entonces $g \circ f$ también es biyectiva. (Demuestra que $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.)

EJERCICIO 3.4 Demuestra que si f y g son transformaciones de un conjunto A entonces $(f \circ g)$ también es una transformación de A.

3.1.1 Grupos de transformaciones

Veremos ahora ciertos ejemplos pequeños de conjuntos de transformaciones, que por su importancia reciben un nombre especial, el de *grupo*.

Consideremos un conjunto con dos elementos, $\{0, 1\}$; llamémoslo Δ_2, que se lee "delta-dos". Las funciones de Δ_2 en sí mismo son cuatro :

$$\begin{array}{ccccccccccccccc} & \text{id} & & & & c_0 & & & & c_1 & & & & \rho & \\ 0 & \longmapsto & 0 & \quad & 0 & \longmapsto & 0 & \quad & 0 & \longmapsto & 1 & \quad & 0 & \longmapsto & 1 \\ 1 & \longmapsto & 1 & \quad & 1 & \longmapsto & 0 & \quad & 1 & \longmapsto & 1 & \quad & 1 & \longmapsto & 0, \end{array}$$

donde hemos usado la notación $x \longmapsto y$ para especificar que el elemento x va a dar al elemento y bajo la función en cuestión (no hay que confundir la flechita con "raya

de salida" $\longmapsto$ con la flecha $\rightarrow$ que denota función; inclusive puede pensarse que esta última ($\rightarrow$) es el conjunto de todas las flechitas ($\longmapsto$) entre los elementos). En nuestro ejemplo, las dos funciones de enmedio son *funciones constantes*; y las únicas transformaciones son las de los extremos, id y ρ, donde hay que observar que ρ es su propio inverso, es decir, $\rho \circ \rho = \text{id}$.

Consideremos ahora a $\Delta_3 := \{0, 1, 2\}$, un conjunto con tres elementos. Una función de Δ_3 en sí mismo puede especificarse en una tablita:

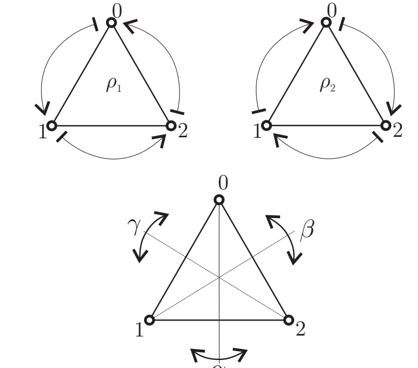
$$\begin{array}{ccc} 0 & \longmapsto & x \\ 1 & \longmapsto & y \\ 2 & \longmapsto & z, \end{array}$$

donde $x, y, z \in \Delta_3$. Como las imágenes $(x, y, z \in \Delta_3)$ son arbitrarias, tenemos que hay $3^3 = 27$ funciones en total; pero de éstas solamente $3 \times 2 \times 1 = 6$ son transformaciones. Pues si queremos que sea biyectiva, una vez que el 0 *escoge* su imagen, al 1 sólo le quedan dos opciones para escoger y cuando lo hace, al 2 ya no le queda más que una opción obligada.

Estas seis transformaciones son:

$$\begin{array}{ccc} & \text{id} & \\ 0 & \longmapsto & 0 \\ 1 & \longmapsto & 1 \\ 2 & \longmapsto & 2 \end{array} \qquad \begin{array}{ccc} & \rho_1 & \\ 0 & \longmapsto & 1 \\ 1 & \longmapsto & 2 \\ 2 & \longmapsto & 0 \end{array} \qquad \begin{array}{ccc} & \rho_2 & \\ 0 & \longmapsto & 2 \\ 1 & \longmapsto & 0 \\ 2 & \longmapsto & 1 \end{array}$$

$$\begin{array}{ccc} & \alpha & \\ 0 & \longmapsto & 0 \\ 1 & \longmapsto & 2 \\ 2 & \longmapsto & 1 \end{array} \qquad \begin{array}{ccc} & \beta & \\ 0 & \longmapsto & 2 \\ 1 & \longmapsto & 1 \\ 2 & \longmapsto & 0 \end{array} \qquad \begin{array}{ccc} & \gamma & \\ 0 & \longmapsto & 1 \\ 1 & \longmapsto & 0 \\ 2 & \longmapsto & 2 \end{array}$$



que podemos visualizar como las "simetrías" de un triángulo equilátero. Vistas así, las tres de arriba corresponden a rotaciones (la identidad rota 0 grados), y cumplen con que ρ_1 y ρ_2 son inversas, es decir $\rho_1 \circ \rho_2 = \rho_2 \circ \rho_1 = \text{id}$, lo que corresponde a que una rota $120°$ en una dirección y la otra $120°$ en la dirección contraria. Pero también cumplen que $\rho_1 \circ \rho_1 = \rho_2$ (que podríamos escribir $\rho_1^2 = \rho_2$ si convenimos en que

$$f^n = \overbrace{f \circ f \circ \ldots \circ f}\,,$$

donde f es cualquier transformación; es decir, f^n es f compuesta consigo misma n veces, lo cual tiene sentido sólo cuando f sale de y llega al mismo conjunto). Por su parte, las tres transformaciones siguientes se llaman *transposiciones* y geométricamente se ven como *reflexiones*. Cumplen que $\alpha^2 = \beta^2 = \gamma^2 = \text{id}$, y además cumplen otras *relaciones* como que $\alpha \circ \beta = \rho_1$, lo cual se ve *persiguiendo* elementos:

$$\begin{array}{ccccc} & \beta & & \alpha & \\ 0 & \longmapsto & 2 & \longmapsto & 1 \\ 1 & \longmapsto & 1 & \longmapsto & 2 \\ 2 & \longmapsto & 0 & \longmapsto & 0; \end{array}$$

o bien, que $\alpha \circ \beta \circ \alpha = \beta \circ \alpha \circ \beta = \gamma$ (¡compruébalo persiguiendo elementos!). Para poder concluir con elegancia, conviene introducir las siguientes nociones generales. Hay ciertos conjuntos de transformaciones que son tan importantes que conviene darles un nombre específico, el de **grupo**:

Definición 3.1.2 A un conjunto G de transformaciones de un conjunto A se le llama un *grupo de transformaciones* de A si cumple

$$\begin{array}{rl} \textbf{i)} & \mathrm{id}_A \in G \\ \textbf{ii)} & f, g \in G \Rightarrow g \circ f \in G \\ \textbf{iii)} & f \in G \Rightarrow f^{-1} \in G. \end{array}$$

Ejemplos triviales de grupos serían el conjunto de *todas* las transformaciones de A, o bien, sólo su identidad, $\{\mathrm{id}_A\}$.

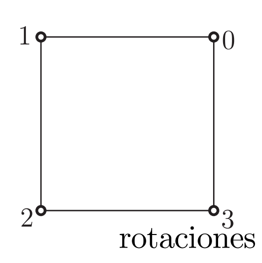
En el caso de $A = \Delta_3$, el grupo de todas sus transformaciones tiene seis elementos; pero también hay otro grupo de transformaciones que consiste en las tres del primer renglón, las rotaciones; pues contiene a la identidad (es un conjunto no vacío), es *cerrado bajo composición* (cumple **(ii)**) y es *cerrado bajo* inversas (cumple **(iii)**). Y también cada una de las transposiciones (o reflexiones) junto con la identidad forman un grupo (con dos elementos) de transformaciones de Δ_3.

Dado un conjunto cualquiera de transformaciones de A, el grupo que *genera* es el grupo de transformaciones que se obtiene de todas las posibles composiciones con elementos de él o sus inversos.

Por ejemplo, α y β generan todas las transformaciones de Δ_3 (pues ya las hemos descrito como composiciones de α y β); mientras que ρ_1 genera el grupo de rotaciones de Δ_3 (que consiste de id, ρ_1 y $\rho_1^2 = \rho_2$). Las *relaciones* que cumplen α y β son

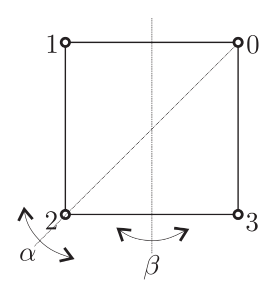
$$\alpha^2 = \beta^2 = (\beta \circ \alpha)^3 = \mathrm{id}\,.$$

Si consideramos ahora a $\Delta_4 := \{0, 1, 2, 3\}$, tendríamos que el conjunto de sus funciones tiene $4^4 = 256$ elementos, mientras que el grupo de todas sus transformaciones tiene $4! = 4 \times 3 \times 2 \times 1 = 24$ elementos (demasiados para escribirlos todos). Pero dentro de éste, podemos encontrar otros grupos, que podemos llamar *subgrupos*.



Uno importante es el de aquellas transformaciones que mantienen intacta (que *preservan*) la estructura del cuadrado regular cuyos vértices se etiquetan $0, 1, 2, 3$ en orden cíclico; por ejemplo, la que mantiene fijos al 0 y al 1 pero que transpone al 2 y al 3 no preserva al cuadrado pues, por ejemplo, la arista del 1 al 2 va a una diagonal, la $1 - 3$. No es dificil ver que éstas (las *simétrias* del cuadrado) son ocho: las cuatro rotaciones (incluyendo a la identidad), y cuatro reflexiones.

Pero en vez de escribirlas todas podemos describirlas mediante *generadores*. Sean ahora



$$\begin{array}{ccc} & \alpha & \\ 0 & \mapsto & 0 \\ 1 & \mapsto & 3 \\ 2 & \mapsto & 2 \\ 3 & \mapsto & 1 \end{array} \qquad \begin{array}{ccc} & \beta & \\ 0 & \mapsto & 1 \\ 1 & \mapsto & 0 \\ 2 & \mapsto & 3 \\ 3 & \mapsto & 2, \end{array} \tag{3.1}$$

entonces $\beta \circ \alpha$ es la rotación $(0 \mapsto 1 \mapsto 2 \mapsto 3 \mapsto 0)$ y las otras dos rotaciones del cuadrado son $(\beta \circ \alpha)^2$ y $(\beta \circ \alpha)^3$; mientras que las dos reflexiones que faltan describir son $(\alpha \circ \beta \circ \alpha)$ y $(\beta \circ \alpha \circ \beta)$. Podemos entonces concluir que el grupo de simetrías del cuadrado está generado por α y β que cumplen las relaciones

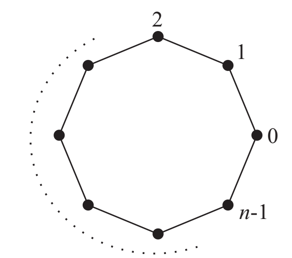
$$\alpha^2 = \beta^2 = (\beta \circ \alpha)^4 = \mathrm{id}\,,$$

mientras que el grupo de rotaciones tiene un solo generador $\rho = (\beta \circ \alpha)$ que cumple $\rho^4 = \mathrm{id}$.

Para referencia posterior, y a reserva de que se estudien con más detenimiento en el caso general en la Sección 3.3.2, conviene poner nombre a los grupos de transformaciones que hemos descrito.

Al conjunto de todas las transformaciones de un conjunto con n elementos $\Delta_n := \{0, 1, ..., n-1\}$ se le llama el *grupo simétrico de orden* n; se le denota $\mathbf{S}_n$ y consta de $n! := n \times (n-1) \times (n-2) \times ... \times 2 \times 1$ (n factorial) elementos también llamados *permutaciones*.

Dentro de este grupo hemos considerado dos subgrupos (para $n = 3, 4$): el que preserva la estructura del polígono regular con n lados cuyos vértices se etiquetan $0, 1, 2, ..., (n-1)$ en orden cíclico, al cual se le llama el *grupo diédrico de orden* n, y se le denota $\mathbf{D}_n$, que tiene $2n$ elementos; y el subgrupo de éste generado por la *rotación* $(0 \mapsto 1 \mapsto 2 \mapsto ... \mapsto (n-1) \mapsto 0)$ que tiene n elementos, se le llama *grupo cíclico de orden* n y se le denota $\mathbf{C}_n$. Hay que observar que para $n = 3$ el diédrico y el simétrico coinciden ($\mathbf{S}_3 = \mathbf{D}_3$) pero esto ya no sucede para $n = 4$.



EJERCICIO 3.5 Para el caso $n = 4$, escribe explícitamente (en forma de tabla de asignaciones) las ocho transformaciones del grupo diédrico $\mathbf{D}_4$; su expresión más económica (en el número de símbolos usados) como composición de α y β (definidas por las tablas de asignaciones (3.1)), y también da su representación geométrica.

EJERCICIO 3.6 ¿Puedes encontrar un subgrupo de $\mathbf{S}_4$ "esencialmente igual" a $\mathbf{S}_3$?

EJERCICIO 3.7 ¿Puedes encontrar un subgrupo de $\mathbf{S}_4$ de orden 12? (Piensa en un tetraedro regular en $\mathbb{R}^3$ y describe el grupo geométricamente.)

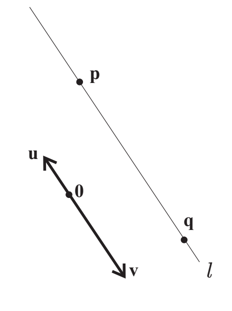
EJERCICIO 3.8 ¿Puedes encontrar una permutacion $\gamma : \Delta_4 \to \Delta_4$ tal que α, β y γ generen todas las permutaciones $\mathbf{S}_4$ (por supuesto que $\gamma \notin \mathbf{D}_4$)?

EJERCICIO 3.9 Da explícitamente (con la tabla de asignación $\Delta_n \to \Delta_n$) dos generadores α y β para los grupos diédricos de orden $n = 5$ y $n = 6$ (es decir, para las simetrías del pentágono y el hexágono). ¿Qué relaciones cumplen? ¿Puedes intuir y describir lo equivalente para el caso general?

***EJERCICIO 3.10** Considera un cubo Q en $\mathbb{R}^3$. Etiqueta sus vértices con Δ_8. Dentro de $\mathbf{S}_8$ hay un subgrupo que es el que preserva la estructura geométrica del cubo. ¿Cuántos elementos tiene? ¿Puedes describirlo con generadores y relaciones?

3.2 Las transformaciones afines de $\mathbb{R}$

El primer grupo de transformaciones geométrico-analíticas que estudiaremos es el que surge de la ambigüedad en la parametrización de rectas. Recuérdese que nuestra definición original de una recta fue con un parametro real, pero para una sola recta hay muchas parametrizaciones (dependen de escoger un punto base y un vector direccional); la forma de relacionarse de estos parametros serán las transformaciones afines.



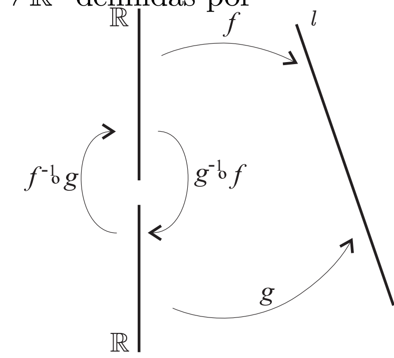
Supongamos que tenemos una misma recta ℓ parametrizada de dos maneras distintas. Es decir, que tenemos

$$\begin{aligned} \ell &= \{\mathbf{p} + t\mathbf{v} : t \in \mathbb{R}\} \\ \ell &= \{\mathbf{q} + s\mathbf{u} : s \in \mathbb{R}\}, \end{aligned}$$

donde $\mathbf{p}$, $\mathbf{q}$, $\mathbf{v}$, $\mathbf{u} \in \mathbb{R}^2$ (aunque también funcione el razonamiento que sigue en cualquier espacio vectorial). Sabemos además que los vectores direccionales, $\mathbf{u}$ y $\mathbf{v}$ son distintos de $\mathbf{0}$ para que efectivamente describan una línea recta. La pregunta es ¿cómo se relacionan los parametros t y s?

Otra manera de pensar estas rectas es como la imagen de funciones cuyo dominio son los reales, es decir, tenemos dos funciones $f, g : \mathbb{R} \to \mathbb{R}^2$ definidas por

$$\begin{aligned} f(t) &= \mathbf{p} + t\mathbf{v} \\ g(s) &= \mathbf{q} + s\mathbf{u}, \end{aligned}$$



cuya imagen es la misma recta ℓ. Puesto que ambas son biyecciones sobre ℓ, podemos restringir el codominio y pensarlas a ambas como funciones de $\mathbb{R}$ en ℓ. Y entonces tiene sentido hablar de sus inversas $f^{-1} : \ell \to \mathbb{R}$ y $g^{-1} : \ell \to \mathbb{R}$. La pregunta es entonces ¿cómo se expresan las funciones $(g^{-1} \circ f)$ y $(f^{-1} \circ g)$ que van de $\mathbb{R}$ en $\mathbb{R}$?

Puesto que $\mathbf{u}$ y $\mathbf{v}$ son paralelos (definen la misma recta), existe un número real $a \in \mathbb{R}$ tal que $\mathbf{u} = a\mathbf{v}$; de hecho $a = (\mathbf{u} \cdot \mathbf{v}) / (\mathbf{v} \cdot \mathbf{v})$, y obsérvese que $a \neq 0$. Y además como $\mathbf{q} \in \ell$, se puede expresar en términos de la primera parametrización; es decir, existe $b \in \mathbb{R}$ tal que $\mathbf{q} = f(b) = \mathbf{p} + b\,\mathbf{v}$. Tenemos entonces que para cualquier $s \in \mathbb{R}$:

$$\begin{aligned} g(s) &= \mathbf{q} + s\,\mathbf{u} \\ &= \mathbf{p} + b\,\mathbf{v} + s(a\,\mathbf{v}) \\ &= \mathbf{p} + (as + b)\,\mathbf{v} \\ &= f(as + b); \end{aligned}$$

al aplicar la función f^{-1} a ambos lados de esta ecuación se obtiene

$$(f^{-1} \circ g)(s) = as + b.$$

Para obtener la otra composición, podemos proceder más directamente, considerando esta última expresión como la que da al parametro t. Es decir

$$t = as + b,$$

de donde, como $a \neq 0$, podemos despejar

$$s = a^{-1}t - ba^{-1}.$$

Definición 3.2.1 Una función $f : \mathbb{R} \to \mathbb{R}$ se llama *afín* si se escribe como

$$f(x) = ax + b, \tag{3.2}$$

donde $a, b \in \mathbb{R}$; y cuando $a \neq 0$ la llamaremos *transformación afín.*

Nuestro uso de el término transformación se justifica por el siguiente ejercicio.

EJERCICIO 3.11 Demuestra que la función afín (3.2) es biyectiva si y sólo si $a \neq 0$.

EJERCICIO 3.12 Sean $f, g : \mathbb{R} \to \mathbb{R}$ definidas por $f(x) = 2x + 1$ y $g(x) = -x + 2$. Encuentra las fórmulas para f^{-1}, g^{-1}, f^2, $f \circ g$ y $g \circ f$.

EJERCICIO 3.13 Demuestra que las transformaciones afines son cerradas bajo inversas y composición.

Obsérvese que las gráficas de las funciones afines son las rectas no verticales, pues éstas están dadas por la ecuación $y = ax+b$. Y que de éstas las rectas no horizontales corresponden a las transformaciones afines.

Al conjunto de todas las transformaciones afines de $\mathbb{R}$, que por el ejercicio anterior forman un grupo, lo denotaremos $\mathbf{Af}(1)$.

Podemos ahora resumir lo que hicimos en los párrafos anteriores.

Lema 3.3 *Dos parametrizaciones de una misma recta se relacionan por una transformación afín entre los parámetros.* □

Así como las parametrizaciones de rectas dependen de la elección de dos puntos en ellas, las transformaciones afines dependen únicamente de dos valores. Pues las constantes que la definen, a y b en nuestro caso (3.2), se obtienen como

$$\begin{aligned} b &= f(0) \\ a &= f(1) - f(0) \end{aligned}$$

y representan un cambio de escala (multiplicar por a) y luego una traslación (sumarle b). De tal manera que si nos dicen que f es una función afín que manda al 0 en b ($f(0) = b$) y al 1 en c ($f(1) = c$), recuperamos toda la función por la fórmula

$$f(x) = (c - b)\,x + b,$$

y obsérvese que es transformación cuando $f(0) \neq f(1)$; si no: es constante. Si el 0 y el 1 pueden ir a cualquier pareja de números por una transformación afín, entonces cualquier pareja de números (distintos) puede ser enviada al 0, 1 (por la inversa) y de ahí a cualquier otra pareja. Hemos demostrado así el:

Teorema 3.4 (Dos en Dos) *Dados dos pares de puntos* x_0, x_1 *y* y_0, y_1 *en* $\mathbb{R}$, *(donde por par se entiende que son distintos,* i.e., $x_0 \neq x_1$ *y* $y_0 \neq y_1$*), existe una única transformación afín* $f \in \boldsymbol{Af}(1)$ *que manda una en la otra,* i.e., *tal que* $f(x_0) = y_0$ *y* $f(x_1) = y_1$. □

EJERCICIO 3.14 Encuentra la transformación afín $f : \mathbb{R} \to \mathbb{R}$ que cumple:

a) $f(2) = 4$ y $f(5) = 1$
b) $f(1) = -1$ y $f(2) = 3$
c) $f(-1) = 0$ y $f(3) = 2$

EJERCICIO 3.15 Encuentra la fórmula explícita (en términos de x_0, x_1, y_0, y_1) de la función f del Teorema 3.4.

EJERCICIO 3.16 Discute qué pasa en el Teorema 3.4 (y en la fórmula del ejercicio anterior) si se permite que en los pares se dé la igualdad.

3.2.1 Isometrías de $\mathbb{R}$

Hemos dicho que las transformaciones afines de la recta consisten de un "cambio de escala" (determinado por la constante a) y luego una traslación (determinada por b). Pero podemos ser más precisos. Como la distancia en $\mathbb{R}$ se mide por la fórmula $d(x, y) = |x - y|$, podemos demostrar que todas las distancias cambian por el mismo factor bajo la transformación afín $f(x) = ax + b$:

$$\begin{aligned} d(f(x), f(y)) &= |ax + b - (ay + b)| = |ax - ay| \\ &= |a(x - y)| = |a|\,|x - y| = |a|\,d(x, y). \end{aligned}$$

Tenemos entonces una familia distinguida de transformaciones afines que son las que **no** cambian la escala, las que mantienen rígida la recta real y que llamaremos *isometrías* de $\mathbb{R}$ porque preservan la métrica (la distancia). Denotemos

$$\mathbf{Iso}(1) := \{f : \mathbb{R} \to \mathbb{R} \mid f(x) = ax + b \quad \text{con} \quad |a| = 1\}$$

(dejamos como ejercicio demostrar que es un grupo de transformaciones).

Tenemos que $\mathbf{Iso}(1)$ se divide naturalmente en dos clases de transformaciones: las *traslaciones*, cuando $a = 1$ (y la función es entonces $f(x) = x + b$), que consisten en deslizar la recta rígidamente hasta que el 0 caiga en b; o bien las *reflexiones*, cuando $a = -1$, y que entonces se escriben $g(x) = -x + b$. Veámos que estas últimas tienen un *punto fijo*, es decir, un punto que se queda en su lugar bajo la transformación. Éste debe satisfacer la ecuación $g(x) = x$ que es

$$x = -x + b$$

y que implica $x = b/2$. Entonces, lo que hacen las reflexiones es intercambiar rígidamente los dos lados de un punto que podemos llamar su *espejo* (pues por ejemplo, $g(b/2 + 1) = b/2 - 1$ y en general se tiene que $g(b/2 + x) = b/2 - x$), y de ahí el nombre de “reflexión”.

EJERCICIO 3.17 Demuestra que $\mathbf{Iso}(1)$ es un grupo de transformaciones de $\mathbb{R}$.

EJERCICIO 3.18 Demuestra que las traslaciones de $\mathbb{R}$, que denotaremos $\mathbf{Tra}(1)$, forman un grupo de transformaciones.

EJERCICIO 3.19 Demuestra que si $f \in \mathbf{Af}(1)$ no tiene puntos fijos (es decir, que $f(x) \neq x$ para toda $x \in \mathbb{R}$), entonces es una traslación no trivial. (La trivial es trasladar por 0, que es la identidad y tiene a todo $\mathbb{R}$ como puntos fijos).

EJERCICIO 3.20 Demuestra, usando la fórmula, que la inversa de cualquier reflexión es ella misma.

EJERCICIO 3.21 Demuestra que la composición de dos reflexiones es una traslación del doble de la distancia (dirigida) entre sus espejos.

3.3 Isometrías y transformaciones ortogonales

En esta sección se estudian las transformaciones más importantes de la geometría euclidiana en su sentido estricto, pues son las que preservan la métrica, la noción de distancia y, por lo tanto, la estructura rígida de las figuras geométricas. Puesto que la definición general es intuitivamente nítida, partiremos de ella y deduciremos las propiedades básicas de las isometrías que nos llevarán, en las secciones siguientes, a obtener las fórmulas explícitas que las definen y que entonces nos facilitarán la obtención de nuevos resultados y su comprensión cabal.

Es importante señalar que muchos de los resultados de esta sección, así como las definiciones, sólo dependen de las nociones de distancia y producto interior en un espacio vectorial. Así que procederemos en general (para $n = 1, 2, 3$, puede pensarse) y hasta el final de la sección, salvo por los ejemplos, concretaremos los resultados al plano.

Definición 3.3.1 Una función $f : \mathbb{R}^n \to \mathbb{R}^n$ es una *isometría* si preserva distancia. Es decir, si para todo par x, $y \in \mathbb{R}^n$ se cumple que

$$d(x, y) = d(f(x), f(y)).$$

También se les llama *transformaciones rígidas* (aunque esta terminología no cobrará sentido hasta que demostremos que son biyectivas). Se denota por $\mathbf{Iso}(n)$ al conjunto de todas las isometrías de $\mathbb{R}^n$.

Nótese que la definición tiene sentido en cualquier espacio métrico, y que coincide para $n = 1$ con las de la sección anterior (3.2.1). Las isometrías mandan al espacio en sí mismo de manera tal que cualquier estructura rígida se mantiene. Estas funciones o transformaciones las vemos a diario. Por ejemplo, mover una silla de un lugar a otro induce una isometría si pensamos que el espacio euclidiano se puede generar y definir respecto de ella; el efecto de esta isometría en la silla es ponerla en su destino.

Antes de ver más ejemplos en detalle, demostraremos tres lemas generales que usan solamente la definición de isometría y tienden hacia la demostración de que las isometrías forman un grupo de transformaciones.

Lema 3.5 *Una isometría* $f : \mathbb{R}^n \to \mathbb{R}^n$ *es inyectiva.*

Demostración. Esto se debe a que la distancia entre puntos diferentes es estrictamente positiva. Formalmente, supongamos que x, $y \in \mathbb{R}^n$ son tales que $f(x) = f(y)$. Esto implica que $d(f(x), f(y)) = 0$. Como f es isometría, entonces $d(x, y) = d(f(x), f(y)) = 0$ y por lo tanto $x = y$. □

Lema 3.6 *Si* f, $g : \mathbb{R}^n \to \mathbb{R}^n$ *son isometrías entonces* $g \circ f$ *también lo es.*

Demostración. Usando la regla de composición y la definición de isometría dos veces (primero para g y luego para f), se obtiene que para cualquier x, $y \in \mathbb{R}^n$:

$$\begin{aligned} d((g \circ f)(x), (g \circ f)(y)) &= d(g(f(x)), g(f(y))) \\ &= d(f(x), f(y)) \\ &= d(x, y) \end{aligned}$$

y por tanto $g \circ f \in \mathbf{Iso}(n)$. □

Lema 3.7 *Si* $f \in \textbf{\textit{Iso}}(n)$ *y tiene inversa* f^{-1}, *entonces* $f^{-1} \in \textbf{\textit{Iso}}(n)$.

Demostración. Dados x, $y \in \mathbb{R}^n$, se tiene que

$$d(f^{-1}(x), f^{-1}(y)) = d(f(f^{-1}(x)), f(f^{-1}(y)))$$

pues f es isometría. Pero la última expresión es $d(x, y)$, lo cual demuestra que $f^{-1} \in \textbf{Iso}(n)$. □

Nos falta entonces demostrar que las isometrías son suprayectivas para concluir que $\textbf{Iso}(n)$ es un grupo de transformaciones. Aunque esto sea intuitivamente claro ("una transformación rígida del plano no puede dejar partes descobijadas"), la demostración sería de momento complicada; con un poco más de técnica será muy sencilla. Así que vale suponer que es cierto por un rato, desarrollar los ejemplos, la intuición y la técnica y luego volver a preocuparnos.

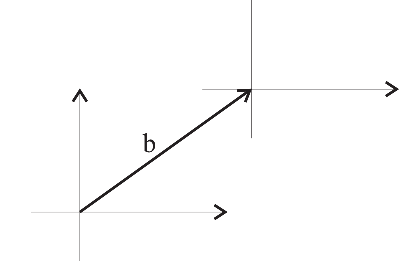
3.3.1 Ejemplos

Ya hemos visto los ejemplos en la recta (cuando $n = 1$), que son traslaciones y reflexiones. Ejemplos en el plano serían rotar alrededor de un punto fijo, reflejar en una recta o bien trasladar por un vector fijo todo el plano. Estas últimas son muy fáciles de definir:

Traslaciones

Dado un vector $\mathbf{b} \in \mathbb{R}^n$, la *traslación por* $\mathbf{b}$ es la función

$$\tau_{\mathbf{b}} : \mathbb{R}^n \to \mathbb{R}^n$$
$$\tau_{\mathbf{b}}(x) = x + \mathbf{b},$$



que claramente es una transformación (inyectiva, pues $x + \mathbf{b} = y + \mathbf{b} \Rightarrow x = y$; y sobre, pues $\tau_{\mathbf{b}}^{-1} = \tau_{-\mathbf{b}}$). Y además es una isometría ($\tau_{\mathbf{b}} \in \textbf{Iso}(n)$) ya que

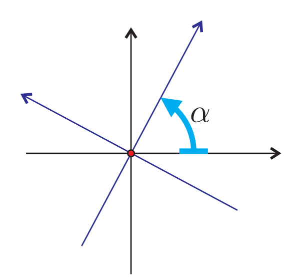
$$\begin{aligned} d(\tau_{\mathbf{b}}(x), \tau_{\mathbf{b}}(y)) &= |\tau_{\mathbf{b}}(x) - \tau_{\mathbf{b}}(y)| \\ &= |(x + \mathbf{b}) - (y + \mathbf{b})| = |x - y| = d(x, y). \end{aligned}$$

De hecho, las traslaciones forman un grupo de transformaciones de $\mathbb{R}^n$, al que denotaremos $\textbf{Tra}(n)$, que puede identificarse con el grupo aditivo $\mathbb{R}^n$; en lenguaje de grupos "son *isomorfos*", pues hay una traslación por cada elemento de $\mathbb{R}^n$, y su regla de composición es claramente

$$\tau_{\mathbf{b}} \circ \tau_{\mathbf{a}} = \tau_{(\mathbf{a}+\mathbf{b})}.$$

Rotaciones

Definir explícitamente las rotaciones es más difícil, aunque intuitivamente sea claro a qué nos referimos: "clavar una tachuela en algún punto y luego rotar el plano alrededor de ella un cierto ángulo". Usando coordenadas polares sí es fácil definir la rotación de un ángulo α alrededor del origen, pues al ángulo de cualquier vector simplemente le sumamos el ángulo de rotación. Así que podemos definir la *rotación de un ángulo* α alrededor del origen como

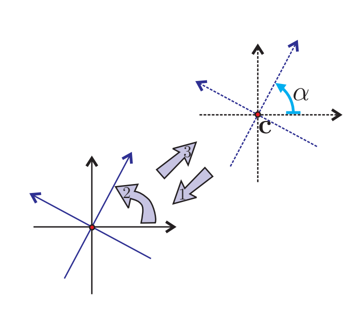


$$\begin{aligned} \rho_\alpha &: \mathbb{R}^2 \to \mathbb{R}^2 \\ \rho_\alpha(\theta, r) &= (\theta + \alpha, r) \quad \text{(en coordenadas polares).} \end{aligned}$$

Nótese que entonces $\rho_0 = \rho_{2\pi} = \mathrm{id}_{\mathbb{R}^2}$ y se cumple, como en las traslaciones, que

$$\rho_\beta \circ \rho_\alpha = \rho_{(\alpha+\beta)},$$

pero ahora sumando ángulos. Además son transformaciones (biyectivas, insistimos en ello) pues tienen inversa $\rho_\alpha^{-1} = \rho_{-\alpha}$ de tal manera que las rotaciones alrededor del origen forman un grupo de transformaciones de $\mathbb{R}^2$ (isomorfo al de los ángulos con la suma: los elementos de este grupo —transformaciones por definición, que más adelante denotaremos con SO (2)— están en correspondencia uno a uno con los puntos del círculo unitario $\mathbb{S}^1$ de tal manera que la composición correponde a la suma de ángulos).



Ahora, usando las traslaciones podemos definir las rotaciones con centro en cualquier otro lado con un truco llamado *conjugación*. Para obtener la rotación de un ángulo α con centro en el punto $\mathbf{c}$, denotémosla con $\rho_{\alpha,\mathbf{c}}$; podemos llevar el centro $\mathbf{c}$ al origen por medio de la traslación $\tau_{-\mathbf{c}}$, rotamos ahí y luego regresamos a $\mathbf{c}$ a su lugar; es decir, podemos definir

$$\rho_{\alpha,\mathbf{c}} = \tau_\mathbf{c} \circ \rho_\alpha \circ \tau_{-\mathbf{c}}.$$

Aunque sea intuitivamente cristalino que las rotaciones son isometrías (se puede rotar un vidrio), no tenemos aún una demostración formal de ello; y con coordenadas polares se ve "en chino" pues expresar distancias (o traslaciones) en términos de ellas está igual de complicado. Mejor nos esperamos a tener buenas expresiones cartesianas de las rotaciones para demostrar que preservan distancias y que al componer dos rotaciones cualesquiera se obtiene otra (el problema por ahora es encontrar su centro). Y entonces veremos que junto con las traslaciones forman el grupo de *movimientos rígidos del plano*, llamados así pues se puede llegar a cualquiera de estas transformaciones *moviendo* continuamente (poco a poco) el plano.

EJERCICIO 3.22 Demuestra formalmente (con las definiciones anteriores) que el conjunto de rotaciones alrededor de un punto dado $\mathbf{c}$ es un grupo.

EJERCICIO 3.23 ¿Puedes encontrar el centro de la rotación de 90° que se obtiene rotando 45° en el origen y luego rotando otros 45° en el punto $(2,0)$? (Haz dibujos —o juega con tachuelas y un acetato sobre un papel cuadriculado— y busca un punto que se quede en su lugar después de las dos rotaciones.) ¿Cuál es el centro de rotación si se invierte el orden de la composición?

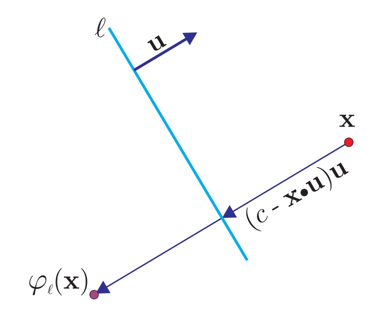
EJERCICIO 3.24 A partir de tu experiencia con el ejercicio anterior, y sin preocuparte por las formalidades, ¿puedes dar una receta geométrica para encontrar el centro de rotación de la composición de dos rotaciones (con centros distintos, por supuesto)?

EJERCICIO 3.25 ¿Puedes conjeturar qué transformación es $\rho_{-\alpha,\mathbf{b}} \circ \rho_{\alpha,\mathbf{c}}$? ¿Puedes demostrarlo? Si no, ¿qué hace falta?

Reflexiones

Las reflexiones (que intercambian rígidamente los dos lados de una "línea espejo") las podemos definir usando la proyección ortogonal a una recta. Dada una recta $\ell \subset \mathbb{R}^2$, se le puede definir por la ecuación normal $\mathbf{u} \cdot \mathbf{x} = c$ con $\mathbf{u}$ un vector unitario y $c \in \mathbb{R}$ una constante (lo que habíamos llamado ecuación unitaria). Sabemos que para cualquier $\mathbf{x} \in \mathbb{R}^2$ el vector que lleva a $\mathbf{x}$ ortogonalmente a la recta ℓ es

$$(c - \mathbf{u} \cdot \mathbf{x})\,\mathbf{u}$$



pues

$$\mathbf{u} \cdot (\mathbf{x} + (c - \mathbf{u} \cdot \mathbf{x})\,\mathbf{u}) = \mathbf{u} \cdot \mathbf{x} + (c - \mathbf{u} \cdot \mathbf{x})\,(\mathbf{u} \cdot \mathbf{u}) = c.$$

Entonces, el reflejado de $\mathbf{x}$ en el *espejo* ℓ será *empujarlo* otro tanto del otro lado de ℓ. Así que definimos la *reflexión* de $\mathbb{R}^2$ a lo largo de $\ell : \mathbf{u} \cdot \mathbf{x} = c$ (con $|\mathbf{u}| = 1$) como

$$\begin{aligned} \varphi_\ell &: \mathbb{R}^2 \to \mathbb{R}^2 \\ \varphi_\ell(\mathbf{x}) &= \mathbf{x} + 2\,(c - \mathbf{u} \cdot \mathbf{x})\,\mathbf{u}. \end{aligned} \tag{3.3}$$

Obsérvese que no nos preocupamos por las propiedades de grupo respecto a las reflexiones porque **no** lo son. Ni siquiera contienen la identidad.

EJERCICIO 3.26 Usando coordenadas (x, y), da fórmulas más sencillas para las reflexiones en los ejes coordenados y verifica que coinciden con la fórmula anterior.

EJERCICIO 3.27 Usando coordenadas da fórmulas sencillas para las reflexiones en las rectas $x = y$ y $x = -y$.

EJERCICIO 3.28 Usando coordenadas da fórmulas sencillas para las reflexiones en las rectas $x = c$ y $y = c$ donde c es cualquier constante.

EJERCICIO 3.29 Define reflexiones en planos de $\mathbb{R}^3$.

EJERCICIO 3.30 Demuestra que las reflexiones son isometrías.

EJERCICIO 3.31 Demuestra que si φ es una reflexión, entonces $\varphi^{-1} = \varphi$.

EJERCICIO 3.32 Demuestra que si φ_ℓ es la reflexión en la recta ℓ, entonces $x \in \mathbb{R}^2$ es un punto fijo de φ_ℓ, *i.e.* $\varphi_\ell(x) = x$, si y sólo si $x \in \ell$.

3.3.2 Grupos de simetrías

Supongamos por un momento (luego lo demostraremos) que las isometrías de $\mathbb{R}^n$ $(n = 1, 2, 3)$ forman un grupo (algo que no es difícil de creer). Entonces podemos definir los "grupos de simetrías" de las figuras, en el plano o de los cuerpos en el espacio, correspondiendo a nuestra intuición milenaria de lo que queremos decir cuando decimos que algo "tiene simetría", por ejemplo al hablar de una flor, de un dibujo o de un edificio.

Para fijar ideas, pensemos en una figura en un papel, y al papel como el plano $\mathbb{R}^2$. La abstracción más elemental es que la figura es un subconjunto F de $\mathbb{R}^2$: los puntos del plano están en F (pintados de negro) o no (de blanco como el papel). Si nos ponemos quisquillosos, podríamos distinguir colores en los puntos (algo que hacen las computadoras hoy día) y tener entonces una función $\gamma : F \to \{\mathtt{Colores}\}$ para ser más precisos. Pero quedémonos con el bulto, una *figura* F es un subconjunto de $\mathbb{R}^n$ (de una vez englobamos los sólidos en $\mathbb{R}^3$). Una *simetría* de F (y hasta el término "sí-metría" lo indica) es una transformación que preserva la métrica y la figura, es decir una $f \in \mathbf{Iso}(n)$ tal que $f(F) = F$ (donde estamos aplicando la transformación a un conjunto para obtener un nuevo conjunto). Podemos definir entonces el *grupo de simetrías* de la figura $F \subset \mathbb{R}^n$ como el conjunto

$$\mathbf{Sim}(F) := \{f \in \mathbf{Iso}(n) \mid f(F) = F\}.$$

Debemos demostrar entonces que además de ser un conjunto de transformaciones de $\mathbb{R}^n$ cumple con las propiedades que lo hacen un grupo; que estamos usando los términos con propiedad:

i) Claramente $\mathrm{id}_{\mathbb{R}^n} \in \mathbf{Sim}(F)$ pues $\mathrm{id}(F) = F$. De tal manera que cualquier figura tiene al menos la simetría trivial; y si la figura F no tiene otra simetría diríamos coloquialmente que **no** tiene simetrías, que es asimétrica.

ii) Supongamos que $f \in \mathbf{Sim}(F)$; entonces al sustituir $F = f(F)$ obtenemos que

$$f^{-1}(F) = f^{-1}(f(F)) = \left(f^{-1} \circ f\right)(F) = \mathrm{id}(F) = F,$$

donde estamos usando nuestra suposición de que $\mathbf{Iso}(n)$ es un grupo y entonces que $f^{-1} \in \mathbf{Iso}(n)$ ya existe.

iii) Si $f,\ g \in \mathbf{Sim}(F)$ entonces, como ninguna de las dos cambia la figura, se obtiene que

$$(f \circ g)(F) = f(g(F)) = f(F) = F.$$

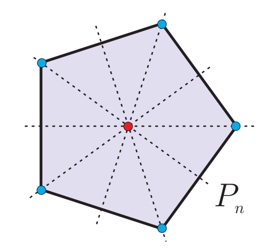
Que dice lo obvio: si un movimiento deja una figura en su lugar le podemos aplicar otro que la deje en su lugar y sigue en su lugar.

De tal manera que cualquier figura o cuerpo tiene asociado un grupo de isometrías; qué tan grande es ese grupo es nuestra medida intuitiva de "su simetría". Por ejemplo, el cuerpo humano (abstracto, por supuesto) tiene como simetría no trivial una reflexión, que al ser su propia inversa forma, junto con la identidad, un grupo. Y ahora sí, podemos definir formalmente los grupos que esbozamos (para $n \geq 5$, y definimos para $n = 3, 4$) en la primera sección de este capítulo.

Grupos diédricos

Consideremos el polígono regular de n lados P_n, con $n \geq 2$, cuyos vértices son el conjunto

$$\{\mathbf{v}_{k,n} := (\cos(2\pi k/n), \operatorname{sen}(2\pi k/n)) \mid k = 0, 1, \cdots, n-1\}$$



y que consiste en los segmentos entre puntos sucesivos $\overline{\mathbf{v}_{k,n}\mathbf{v}_{k+1,n}}$ llamados sus *caras*, *lados* o *aristas*; y si se quiere se puede pensar que también lo de adentro es parte de P_n. Definimos entonces el *grupo diédrico de orden* n como su grupo de simetrías:

$$\mathbf{D}_n := \mathbf{Sim}(P_n).$$

Es claro que $\mathbf{D}_n$ contiene las n rotaciones en el origen $\rho_{2\pi k/n}$ que incluyen la identidad ($k = 0$ o $k = n$). Pero además tiene reflexiones cuyos espejos son las líneas por el origen y los vértices y las líneas por el origen y el punto medio de sus lados (que para n impar ya estaban contados). En cualquier caso, son exactamente n reflexiones pues en nuestro conteo hubo repeticiones. De tal manera que $\mathbf{D}_n$ tiene $2n$ elementos.

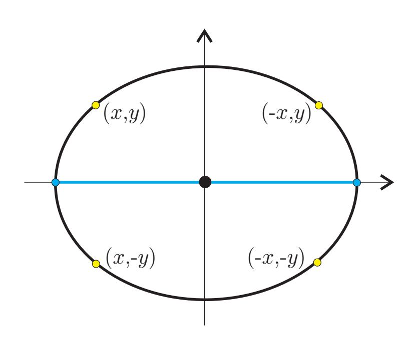
Esta clase de simetría es la que tienen los rosetones de las iglesias (por lo que también se les conoce como "grupos de rosetas o bien de rosetones"); muchas de las figuras de papel picado que se obtienen al doblar radialmente y luego cortar con tijeras; algunas plazas y edificios famosos y múltiples flores así como las estrellas de mar.

Más cercano al "corazoncito" de este libro, podemos considerar las cónicas. Sea $\mathcal{E}$ la elipse dada por la ecuación canónica

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

con $a > b$ (aunque lo que importa es que $a \neq b$). Es claro por el dibujo que las simetrías de $\mathcal{E}$ son las reflexiones en los ejes. Pero lo podemos demostrar formalmente pues, como se debió haber respondido al ejercicio correspondiente, la reflexión en el eje x está dada por la fórmula

$$\varphi_x(x, y) = (x, -y).$$



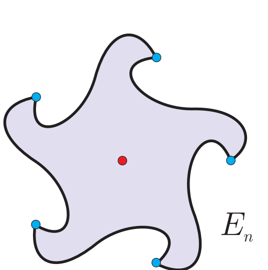
Y análogamente, la reflexión en el eje y es $\varphi_y(x, y) = (-x, y)$, de tal manera que su composición (en cualquier orden) es la rotación de π en el origen ($(x, y) \mapsto (-x, -y)$). Si el punto (x, y) está en la elipse, satisface la ecuación y entonces los otros tres puntos que se obtienen al cambiar el signo a una o a las dos coordenadas también la satisfacen pues cambiar el signo no afecta los cuadrados. Esto implica que $\mathcal{E}$ va en sí misma al reflejar en cualquiera de los ejes, que en adelante llamaremos sus *ejes de simetría.* Falta ver que no tiene ninguna otra simetría y esto se sigue de que su *diámetro* (el segmento máximo entre sus puntos, que mide $2a$) es único. Y puesto que $\mathbf{D}_2$ es, por definición, el grupo de simetrías de un segmento, entonces

$$\mathbf{Sim}(\mathcal{E}) = \mathbf{D}_2.$$

Análogamente, el grupo de simetrías de una hipérbola es el diédrico de orden 2. Las parábolas tienen menos simetrías, sólo una reflexión, que junto con la identidad forma un grupo que podemos llamar $\mathbf{D}_1$ para completar la definición de los diédricos (finitos). Nótese que este tipo de simetría (el de una sola reflexión, junto con la identidad) es muy común, el cuerpo humano (en abstracto, por supuesto), casi todos los animales, una silla, etcétera.

Grupos cíclicos

Son el subgrupo del diédrico correspondiente que consiste en tomar únicamente las rotaciones. Pero lo importante es que también son grupos de simetría de figuras.



Para obtener una, podemos reemplazar las aristas del polígono P_n por curvas que distingan entre salida y llegada, por ejemplo, un símbolo de "integral" para obtener a la "estrella ninja de n picos" E_n. Y entonces el *grupo cíclico de orden* n es

$$\mathbf{C}_n := \mathbf{Sim}(E_n).$$

El grupo cíclico de orden n, $\mathbf{C}_n$, tiene n elementos. Se puede identificar con el grupo de residuos módulo n, denotado con $\mathbb{Z}_n$, donde la composición corresponde a la suma.

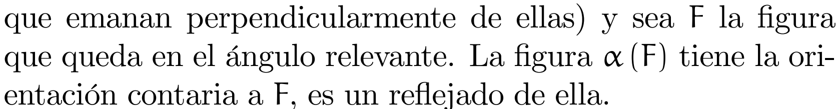
Parece que el gran Leonardo da Vinci notó que cualquier figura plana que se pueda dibujar en una hoja de papel (traducida a una figura $F \subset \mathbb{R}^2$ acotada) y que no tenga todas las simetrías de un círculo (que no sea un conjunto de anillos concéntricos) tiene como grupo de simetrías a alguno de estos dos tipos; es decir, o es diédrico o es cíclico. Una versión de este resultado la demostraremos hacía el final del capítulo, aunque desde ahora se puede entender su enunciado. Dice que si G es un grupo de isometrías que, como conjunto, es finito, entonces es cíclico o es diédrico; que con estos dos ejemplos ya acabamos.

Sin embargo, hay otros grupos de simetría interesantes. Si pensamos en figuras no acotadas; por ejemplo, una cuadrícula o un mosaico (un cubrimiento del plano con copias de unas cuantas piezas), M digamos, entonces $\mathbf{Sim}(M)$ es lo que se llama un grupo cristalográfico. De estos solamente hay 17, pero no lo demostraremos en este libro.

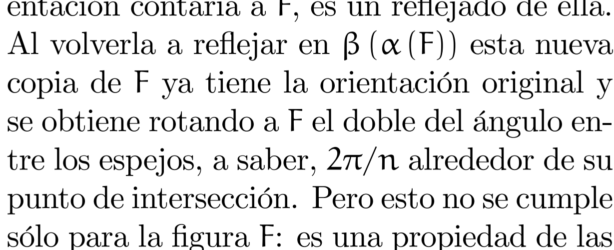
Diédrico y cíclico infinitos

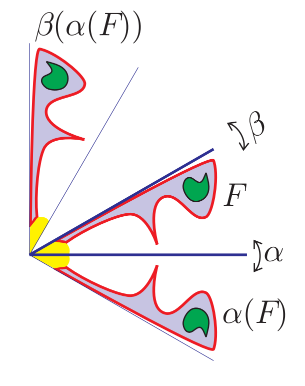
Para entender el porqué del nombre en sus versiones infinitas, vale la pena retomar la idea de generadores de un grupo usada en la primera sección. Para generar una figura con simetría diédrica (que quizá, como veremos, debía llamarse *caleidoscópica*) basta poner dos espejos en un ángulo π/n y colocarlos sobre cualquier dibujo; la porción de éste que queda entre ellos se "reproduce" caleidoscópicamente y forma alrededor de la arista donde se juntan los espejos una figura con simetría $\mathbf{D}_n$.

Consideremos esta situación en el plano del dibujo. Sean α y β las reflexiones en los dos espejos (que ahora son líneas en el plano y los espejos verdaderos son los planos que emanan perpendicularmente de ellas) y sea F la figura que queda en el ángulo relevante. La figura $\alpha(F)$ tiene la orientación contaria a F, es un reflejado de ella.

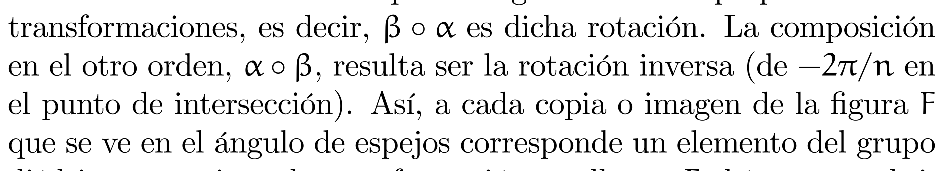


Al volverla a reflejar en $\beta(\alpha(F))$ esta nueva copia de F ya tiene la orientación original y se obtiene rotando a F el doble del ángulo entre los espejos, a saber, $2\pi/n$ alrededor de su punto de intersección.

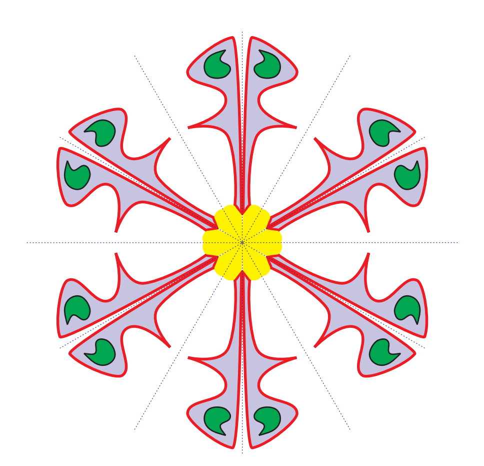




Pero esto no se cumple sólo para la figura F: es una propiedad de las transformaciones, es decir, $\beta \circ \alpha$ es dicha rotación. La composición en el otro orden, $\alpha \circ \beta$, resulta ser la rotación inversa (de $-2\pi/n$ en el punto de intersección). Así, a cada copia o imagen de la figura F que se ve en el ángulo de espejos corresponde un elemento del grupo diédrico que es justo la transformación que lleva a F ahí y que se obtiene componiendo sucesivamente las dos reflexiones originales α y β.



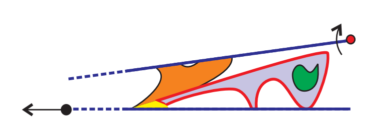
La figura con simetría diédrica o caleidoscópica que se genera es la unión de todas estas copias (justo $2n$) de lo que se podría llamar la *figura fundamental* que habíamos denotado F.

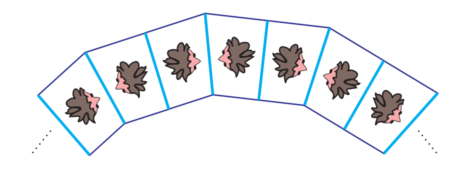


Hemos visto que, como en los casos $n = 3, 4$, los grupos diédricos están generados por dos reflexiones; a saber, $\mathbf{D}_n$ está generado por dos reflexiones cuyos espejos se encuentran en un ángulo π/n (en las figuras tomamos $n = 6$). ¿Qué pasa si hacemos crecer n? Es claro que la figura fundamental F tenderá a hacerse más flaca y astillada, mientras que la figura caleidoscópica que generan (F y las reflexiones α y β), que podemos denotar $\mathbf{D}_n(F)$, pues resulta natural definir

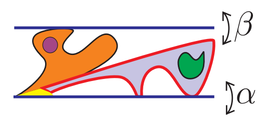
$$\mathbf{D}_n(F) := \bigcup_{g \in \mathbf{D}_n} g\,(F)\,,$$

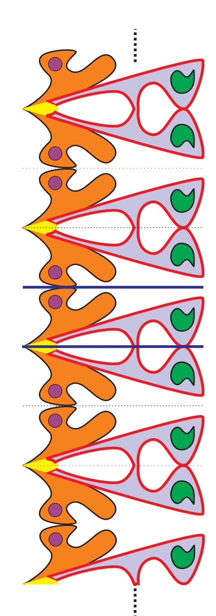
se hará más pesada y confusa, pues el ángulo π/n disminuye. Sin embargo, podemos evitar este tumulto si al mismo tiempo que n crece y π/n disminuye, alejamos el punto de intersección de los espejos; es decir, podemos girar uno de los espejos (el correspondiente a β, digamos) alrededor de algún punto fijo y esto produce que el punto de intersección se aleje. La figura $\mathbf{D}_n(F)$ que se genera entonces es un anillo cuyo radio es grande: es como esos pasillos circulares que se generan en un baño o closet con espejos encontrados pero no perfectamente paralelos; aunque los espejos se acaben, los planos que definen se intersectan en una línea (imaginaria) que es el eje, el centro, del pasillo circular donde están nuestras imágenes viendo alternadamente en ambos sentidos y girando poco a poco.





Desde luego, este proceso tiene un limite: cuando n tiende a infinito el espejo de β llega sin problemas a una paralela al espejo de α; y en ese momento generan el *grupo diédrico infinito* $\mathbf{D}_\infty$. Es el grupo que aparece en las peluquerías donde ponen espejos en paredes opuestas y, en teoría, paralelas. Cada elemento de este grupo es una composición (*palabra*) del estilo $\alpha\beta\alpha\beta\cdots\beta\alpha$, aunque puede empezar y terminar con cualesquiera de α o β. El subgrupo que consiste de palabras pares (incluyendo a la vacía que representa la identidad), que son composición de un número par de reflexiones, es el *cíclico infinito*, $\mathbf{C}_\infty$: consta de puras traslaciones y está generado por la más chica de ellas: $\alpha\beta$ o $\beta\alpha$, da lo mismo. De tal manera que naturalmente se identifica con los enteros ($\mathbb{Z}$) y la suma corresponde a la composición.





Pero el grupo diédrico infinito, como la figura con su simetría lo indica, realmente vive en la recta. Cualquier par de reflexiones distintas en $\mathbf{Iso}(1)$ lo generan. Y si lo definimos como subgrupo de isometrías de $\mathbb{R}^2$, fue porque gráficamente es más fácil de ver y para justificar la naturalidad de su pomposo nombre.

Nótese, por último, que aunque al escoger distintas reflexiones para generar el diédrico infinito (o cualquiera de los finitos, para este caso) se pueden obtener subconjuntos (estrictamente) distintos de isometrías, pero la estructura algebraica de ellos (cómo se comportan bajo composición) es esencialmente la misma. De tal manera que los grupos diédricos y cíclicos que hemos definido deben pensarse como algo abstracto que tiene muchas instancias de realización geométrica. Así que, estrictamente hablando, cuando decimos que el grupo generado por tal par de reflexiones es **el** diédrico fulano, realmente deberíamos decir "es **un** diédrico de zutanito orden".

EJERCICIO 3.33 Sean α y β las dos reflexiones de $\mathbb{R}$ con espejos en el 0 y en el 1 respectivamente, y sea (para este y los siguientes ejercicios) $\mathbf{D}_\infty \subset \mathbf{Iso}(1)$ el grupo de transformaciones que generan α y β. Da las fórmulas explícitas de α, β, $\alpha\circ\beta$, $\beta\circ\alpha$ y $\alpha\circ\beta\circ\alpha$.

EJERCICIO 3.34 ¿Cuál es el grupo $\mathbf{D}_\infty \cap \mathbf{Tra}(1)$? Da la fórmula explícita de todos sus elementos y su correspondiente expresión como generado por α y β. (*El próximo ejercicio te puede ayudar.*)

EJERCICIO 3.35 Demuestra que las reflexiones en $\mathbf{D}_\infty$ son precisamente las reflexiones con espejo entero. Es decir, cuyo espejo está en $\mathbb{Z} = \{... -2, -1, 0, 1, 2, ...\} \subset \mathbb{R}$.

EJERCICIO 3.36 ¿Cuáles son los generadores del grupo diédrico infinito cuyo subgrupo de traslaciones corresponde precisamente a las traslaciones por números enteros $\mathbb{Z}$.

3.3.3 Transformaciones ortogonales

Hemos definido isometrías, visto los ejemplos básicos y desarrollado la importante noción de simetría —y no sólo importante en las matemáticas sino en la naturaleza donde parece ser muy abundante—, pero tenemos aún muy pocas herramientas técnicas para demostrar cosas tan elementales como que $\mathbf{Iso}(n)$ es un grupo. Para esto serán fundamentales las transformaciones ortogonales. Pero su motivación más natural viene de recapitular cómo introdujimos la noción de distancia: como una consecuencia natural del producto interior que era más elemental y fácil de trabajar; y a éste lo hemos tenido olvidado. Si las isometrías se definieron como las funciones que preservan distancia, algo similar se debe poder hacer con el producto interior.

Definición 3.3.2 Una función $f : \mathbb{R}^n \to \mathbb{R}^n$ es una *transformación ortogonal* si preserva el producto interior. Es decir, si para todo par $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ se cumple que

$$\mathbf{x} \cdot \mathbf{y} = f(\mathbf{x}) \cdot f(\mathbf{y}).$$

Se denota por $\mathbf{O}(n)$ el grupo de *transformaciónes ortogonales* de $\mathbb{R}^n$.

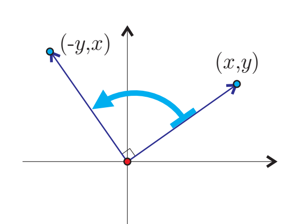
El lector observador debió haber notado que nos estamos adelantando al usar los términos “transformación” y “grupo” en la definición anterior. Estrictamente hablando, deberíamos usar los términos más modestos de “función” y “conjunto”, pero como sé (autor) que sí va a tener sentido, vale la pena irse acostumbrando; además, en este caso me suenan horrible los términos modestos , no me hallo usándolos. Por lo pronto, para limpiar un poco la conciencia, dejamos al lector quisquilloso que cumpla los primeros trámites hacía la “liberación total” de los términos verdaderos.

EJERCICIO 3.37 Demuestra que la composición de funciones ortogonales es ortogonal.

EJERCICIO 3.38 Demuestra que si una función ortogonal tiene inversa entonces ésta también es ortogonal.

EJERCICIO 3.39 Demuestra que $\mathbf{O}(1) = \{\mathrm{id}_{\mathbb{R}}, -\mathrm{id}_{\mathbb{R}}\}$, donde $(-\mathrm{id}_{\mathbb{R}})(x) = -x$. Recuerda que el producto interior en $\mathbb{R}$ es el simple producto. Observa que entonces las transformaciones ortogonales de $\mathbb{R}$ son sus isometrías que dejan fijo el origen.

Antes de entrar a las demostraciones, veamos un ejemplo que nos ha sido muy útil, el compadre ortogonal en $\mathbb{R}^2$. Hemos usado al compadre ortogonal de un vector dado, pero, como lo indicamos en el momento de su presentación, si pensamos en todas las asignaciones al mismo tiempo obtenemos una función de $\mathbb{R}^2$ en $\mathbb{R}^2$, que podemos designar como $\rho : \mathbb{R}^2 \to \mathbb{R}^2$ y que está dada por la fórmula explícita



$$\rho(x, y) = (-y, x).$$

Para demostrar que el compadre ortogonal es una transformación ortogonal (valga y explíquese la redundancia), hay que escribir la fórmula para $\rho(\mathbf{x}) \cdot \rho(\mathbf{y})$ con (conviene cambiar a subíndices) $\mathbf{x} = (x_1, x_2)$ y $\mathbf{y} = (y_1, y_2)$ cualquier par de vectores:

$$\begin{aligned} \rho(\mathbf{x}) \cdot \rho(\mathbf{y}) &= (-x_2, x_1) \cdot (-y_2, y_1) \\ &= (-x_2)(-y_2) + x_1 y_1 \\ &= x_1 y_1 + x_2 y_2 = \mathbf{x} \cdot \mathbf{y} \end{aligned}$$

(de hecho esto fue el inciso **(iii)** del Lema 1.12 que presentaba sus monerías en sociedad). Si el compadre ortogonal que, como función, consiste en girar 90° alrededor del origen es ortogonal, es de esperarse que todas las rotaciones en el origen también lo sean, y así será. Pero antes, relacionemos las transformaciones ortogonales con las isometrías.

Puesto que la distancia se define en términos del producto interior, debe cumplirse que una función ortogonal también es isometría.

Lema 3.8 $\boldsymbol{O}(n) \subset \boldsymbol{Iso}(n)$.

Demostración. Sea $f \in \mathbf{O}(n)$. Debemos demostrar que preserva distancias sabiendo que preserva producto interior. Pero esto es fácil pues la distancia se escribe en términos del producto interior. Dados $\mathbf{x}$ y $\mathbf{y}$ en $\mathbb{R}^n$, la definición de distancia es

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})}.$$

Al elevar al cuadrado y usar las propiedades del producto interior se tiene entonces

$$d(\mathbf{x}, \mathbf{y})^2 = \mathbf{x} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} - 2\mathbf{x} \cdot \mathbf{y}. \tag{3.4}$$

Como esto se cumple para cualquier par de puntos, en particular para $f(\mathbf{x})$ y $f(\mathbf{y})$, tenemos

$$\begin{aligned} d(f(\mathbf{x}), f(\mathbf{y}))^2 &= f(\mathbf{x}) \cdot f(\mathbf{x}) + f(\mathbf{y}) \cdot f(\mathbf{y}) - 2f(\mathbf{x}) \cdot f(\mathbf{y}) \\ &= \mathbf{x} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} - 2\mathbf{x} \cdot \mathbf{y} \\ &= d(\mathbf{x}, \mathbf{y})^2, \end{aligned}$$

donde usamos que $f \in \mathbf{O}(n)$ en la segunda igualdad. Puesto que las distancias son positivas, de aquí se sigue que

$$d(f(\mathbf{x}), f(\mathbf{y})) = d(\mathbf{x}, \mathbf{y}),$$

y por tanto que $f \in \mathbf{Iso}(n)$. □

EJERCICIO 3.40 Demuestra que las funciones ortogonales son inyectivas.

Ahora veremos que lo único que le falta a una isometría para ser ortogonal es respetar el origen. Para que tenga sentido, hay que remarcar que una función ortogonal $f \in \mathbf{O}(n)$ preserva el origen pues éste es el único vector cuyo producto interior consigo mismo se anula, es decir, $f(\mathbf{0}) \cdot f(\mathbf{0}) = \mathbf{0} \cdot \mathbf{0} = 0 \quad \Rightarrow \quad f(\mathbf{0}) = \mathbf{0}$.

Lema 3.9 *Sea* $f \in \textbf{\textit{Iso}}(n)$ *tal que* $f(\mathbf{0}) = \mathbf{0}$*, entonces* $f \in \textbf{\textit{O}}(n)$*.*

Demostración. Con la misma idea del lema anterior, hay que escribir ahora el producto interior en términos de distancias. De la ecuación (3.4), se sigue

$$\mathbf{x} \cdot \mathbf{y} = \frac{1}{2}(\mathbf{x} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} - d(\mathbf{x}, \mathbf{y})^2).$$

Y como $\mathbf{x} \cdot \mathbf{x} = d(\mathbf{x}, \mathbf{0})^2$, tenemos entonces

$$\mathbf{x} \cdot \mathbf{y} = \frac{1}{2}(d(\mathbf{x}, \mathbf{0})^2 + d(\mathbf{y}, \mathbf{0})^2 - d(\mathbf{x}, \mathbf{y})^2)$$

y ya estamos armados para enfrentar el lema.

Sea $f \in \mathbf{Iso}(n)$ (preserva distancias) tal que $f(\mathbf{0}) = \mathbf{0}$. Por lo anterior, se obtiene

$$\begin{aligned} f(\mathbf{x}) \cdot f(\mathbf{y}) &= \frac{1}{2}(d(f(\mathbf{x}), \mathbf{0})^2 + d(f(\mathbf{y}), \mathbf{0})^2 - d(f(\mathbf{x}), f(\mathbf{y}))^2) \\ &= \frac{1}{2}(d(f(\mathbf{x}), f(\mathbf{0}))^2 + d(f(\mathbf{y}), f(\mathbf{0}))^2 - d(f(\mathbf{x}), f(\mathbf{y}))^2) \\ &= \frac{1}{2}(d(\mathbf{x}, \mathbf{0})^2 + d(\mathbf{y}, \mathbf{0})^2 - d(\mathbf{x}, \mathbf{y})^2) \\ &= \mathbf{x} \cdot \mathbf{y}. \end{aligned}$$

□

Estamos ya en posición de expresar todas las isometrías en términos de las ortogonales y las traslaciones.

Proposicion 3.10 *Sea* $f \in \textbf{\textit{Iso}}(n)$*. Entonces existen* $g \in \textbf{\textit{O}}(n)$ *y* $\mathbf{b} \in \mathbb{R}^n$ *tales que para toda* $\mathbf{x} \in \mathbb{R}^n$

$$f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{b},$$

y además son únicas.

Dibujo

Demostración. Veamos primero la unicidad, es decir, que si suponemos que existen $g \in \mathbf{O}(n)$ y $\mathbf{b} \in \mathbb{R}^n$ tales que f se escribe $f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{b}$, entonces g y $\mathbf{b}$ están forzadas. Puesto que las transformaciones ortogonales dejan fijo el origen, obtenemos que

$$f(\mathbf{0}) = g(\mathbf{0}) + \mathbf{b} = \mathbf{0} + \mathbf{b} = \mathbf{b},$$

y entonces la constante $\mathbf{b}$ está forzada a ser $\mathbf{b} := f(0)$. Pero entonces g ya queda definida por la ecuación

$$g(\mathbf{x}) = f(\mathbf{x}) - \mathbf{b}$$

para toda $\mathbf{x} \in \mathbb{R}^n$. Nos falta entonces demostrar que con esta última definición de la función g y el vector constante $\mathbf{b} \in \mathbb{R}^n$ se cumple que g es ortogonal.

Observése que otra manera de definir a g sería $g = \tau_{-\mathbf{b}} \circ f$, donde hay que recordar que $\tau_{-\mathbf{b}}$ es la traslación por $-\mathbf{b}$. Por el Lema 3.6 (la composición de isometrías es isometría) y puesto que las traslaciones son isometrías, se obtiene que $g \in \mathbf{Iso}(n)$. Pero además, como

$$g(0) = f(0) - \mathbf{b} = \mathbf{b} - \mathbf{b} = 0,$$

entonces $g \in \mathbf{O}(n)$ por el lema anterior y la existencia queda demostrada. □

Tenemos entonces que las isometrías se obtienen de las transformaciones ortogonales al permitir que interactúen, que se involucren, que se compongan, con las traslaciones (que **no** son ortogonales, de hecho sólo se intersectan en la identidad, $\mathbf{O}(n) \cap \mathbf{Tra}(n) = \{\mathrm{id}\}$, pues la única traslación que deja fijo el origen es por 0). Corresponde esto al hecho de que la distancia se obtiene del producto interior trasladando un punto al origen. Y por tanto, para acabar de entender $\mathbf{Iso}(n)$ debemos estudiar $\mathbf{O}(n)$. Como en un ejercicio ya establecimos el caso de $\mathbf{O}(1) \simeq \{1, -1\}$, concentrémonos en $\mathbf{O}(2)$.

Sea $f \in \mathbf{O}(2)$. Como f preserva el producto interior, también preserva la norma y en particular manda al círculo unitario $\mathbb{S}^1$ en sí mismo:

$$\mathbf{x} \in \mathbb{S}^1 \Rightarrow \mathbf{x} \cdot \mathbf{x} = 1 \Rightarrow f(\mathbf{x}) \cdot f(\mathbf{x}) = 1 \Rightarrow f(\mathbf{x}) \in \mathbb{S}^1.$$

Pero también preserva ortogonalidad (definida con base en el producto interior). Entonces debe mandar bases ortonormales en bases ortonormales. Veámoslo con el ejemplo más simple. Dibujo

Sean $\mathbf{e}_1$, $\mathbf{e}_2$ la base canónica de $\mathbb{R}^2$; es decir $\mathbf{e}_1 = (1, 0)$ y $\mathbf{e}_2 = (0, 1)$. Y sean

$$\mathbf{u} := f(\mathbf{e}_1)$$
$$\mathbf{v} := f(\mathbf{e}_2).$$

Afirmamos que $\mathbf{u}$ y $\mathbf{v}$ son una base ortonormal, pues de que $f \in \mathbf{O}(2)$ (preserva producto interior) se obtiene

$$\mathbf{u} \cdot \mathbf{u} = \mathbf{e}_1 \cdot \mathbf{e}_1 = 1$$
$$\mathbf{v} \cdot \mathbf{v} = \mathbf{e}_2 \cdot \mathbf{e}_2 = 1$$
$$\mathbf{u} \cdot \mathbf{v} = \mathbf{e}_1 \cdot \mathbf{e}_2 = 0.$$

Y entonces podemos encontrar el valor de $f(\mathbf{x})$ para cualquier $\mathbf{x}$ usando el teorema de las bases ortonormales (Teorema 1.23) sobre $f(\mathbf{x})$:

$$\begin{aligned} f(\mathbf{x}) &= (f(\mathbf{x}) \cdot \mathbf{u})\mathbf{u} + (f(\mathbf{x}) \cdot \mathbf{v})\mathbf{v} \\ &= (f(\mathbf{x}) \cdot f(\mathbf{e}_1))\mathbf{u} + (f(\mathbf{x}) \cdot f(\mathbf{e}_2))\mathbf{v} \\ &= (\mathbf{x} \cdot \mathbf{e}_1)\mathbf{u} + (\mathbf{x} \cdot \mathbf{e}_2)\mathbf{v}, \end{aligned}$$

Dibujo

que se puede escribir, tomando como de costumbre $\mathbf{x} = (x, y)$, como

$$f(x, y) = x\,\mathbf{u} + y\,\mathbf{v}. \tag{3.5}$$

Hemos llegado entonces a una fórmula explícita para las funciones ortogonales (de $\mathbb{R}^2$) que depende únicamente de sus valores en la base canónica. Por ejemplo, si f fuera la función "compadre ortogonal", entonces $\mathbf{u} = f(\mathbf{e}_1) = \mathbf{e}_2$ y $\mathbf{v} = f(\mathbf{e}_2) = -\mathbf{e}_1$ y por lo tanto $f(x, y) = x\,\mathbf{e}_2 - y\,\mathbf{e}_1 = (-y, x)$ como ya sabíamos. Podemos resumir con la:

Proposicion 3.11 *Si $f \in \boldsymbol{O}(2)$, existe una base ortonormal $\mathbf{u}$, $\mathbf{v}$ de $\mathbb{R}^2$ para la cual f se escribe*

$$f(x, y) = x\,\mathbf{u} + y\,\mathbf{v}.$$

De aquí podemos sacar por el momento dos resultados: el primero, que estas funciones ortogonales son suprayectivas (y por lo tanto que ya les podemos decir "transformaciones" y a $\mathbf{O}(2)$ "grupo" con todas las de la ley); y el segundo, que es una descripción geométrica de todas ellas.

Dibujo

Primero, con la notación de arriba, dado cualquier $\mathbf{y} \in \mathbb{R}^2$, es fácil encontrar $\mathbf{x}$ tal que $f(\mathbf{x}) = \mathbf{y}$. Sea $\mathbf{x} := (\mathbf{y} \cdot \mathbf{u})\mathbf{e}_1 + (\mathbf{y} \cdot \mathbf{v})\mathbf{e}_2$, entonces

$$\begin{aligned} f(\mathbf{x}) &= (\mathbf{x} \cdot \mathbf{e}_1)\mathbf{u} + (\mathbf{x} \cdot \mathbf{e}_2)\mathbf{v} \\ &= (\mathbf{y} \cdot \mathbf{u})\mathbf{u} + (\mathbf{y} \cdot \mathbf{v})\mathbf{v} = \mathbf{y} \end{aligned}$$

por el Teorema 1.23, y esto demuestra que f es sobre.

Segundo, cuántas transformaciones (¡ahora sí! sin pruritos) ortogonales de $\mathbb{R}^2$ hay es equivalente a cuántas bases ortonormales podemos escoger. El primer vector $\mathbf{u}$ (la imagen de $\mathbf{e}_1$, insistamos) puede ser cualquiera en el círculo unitario $\mathbb{S}^1$; pero una vez escogido ya nada más nos quedan dos posibilidades para escoger el segundo, pues éste debe ser perpendicular y unitario. Puede ser el compadre ortogonal ($\mathbf{v} = \mathbf{u}^\perp$) o bien su negativo ($\mathbf{v} = -\mathbf{u}^\perp$). En el primer caso, la transformación resulta ser una rotación que manda a la base canónica $\mathbf{e}_1$, $\mathbf{e}_2$ en la base rotada $\mathbf{u}$, $\mathbf{u}^\perp$. Y en el segundo caso veremos que es una reflexión. Podríamos decir entonces que $\mathbf{O}(2)$ "consiste en dos copias" de $\mathbb{S}^1$ parametrizadas por $\mathbf{u}$: una corresponde a las rotaciones (cuando se escoje $\mathbf{v} = \mathbf{u}^\perp$) y la otra a las reflexiones.

Pero para que todo esto sea un hecho nos falta demostrar que sin importar que base ortonormal tomemos, siempre se obtiene una transformación ortogonal.

Proposicion 3.12 *Sea* $\mathbf{u}$, $\mathbf{v}$ *una base ortonormal de* $\mathbb{R}^2$. *Entonces la función* $f : \mathbb{R}^2 \to \mathbb{R}^2$, *definida por*

$$f(x, y) = x\,\mathbf{u} + y\,\mathbf{v},$$

es una transformación ortogonal.

Demostración. Sean $\mathbf{x} = (x_1, x_2)$ y $\mathbf{y} = (y_1, y_2)$ cualquier par de puntos en $\mathbb{R}^2$. Entonces

$$\begin{aligned} f(\mathbf{x}) \cdot f(\mathbf{y}) &= (x_1\,\mathbf{u} + x_2\,\mathbf{v}) \cdot (y_1\,\mathbf{u} + y_2\,\mathbf{v}) \\ &= (x_1 y_1)\,\mathbf{u} \cdot \mathbf{u} + (x_2 y_2)\,\mathbf{v} \cdot \mathbf{v} + (x_1 y_2 + x_2 y_1)\,\mathbf{u} \cdot \mathbf{v} \\ &= x_1 y_1 + x_2 y_2 = \mathbf{x} \cdot \mathbf{y}. \end{aligned}$$

□

EJERCICIO 3.41 Demuestra que las isometrías de $\mathbb{R}^2$ son suprayectivas, para concluir con la demostración de que **Iso**(2) es un grupo de transformaciones.

EJERCICIO 3.42 Demuestra que las funciones ortogonales de $\mathbb{R}^3$ son suprayectivas y concluye las demostraciones de que **O**(3) e **Iso**(3) son grupos de transformaciones.

EJERCICIO 3.43 Sea $f \in \mathbf{O}(2)$ con base ortonormal asociada $\mathbf{u}$, $\mathbf{v}$ (es decir, con $\mathbf{u} = f(\mathbf{e}_1)$ y $\mathbf{v} = f(\mathbf{e}_2)$). Con nuestra demostración de la suprayectividad de f se puede encontrar la base ortonormal asociada a la inversa de f, llamémosla $\mathbf{u}'$, $\mathbf{v}'$. Da sus coordenadas explícitas tomando $\mathbf{u} = (u_1, u_2)$ y $\mathbf{v} = (v_1, v_2)$.

EJERCICIO 3.44 Sean $f \in \mathbf{O}(2)$ y ℓ una recta dada por la ecuación $\mathbf{n} \cdot \mathbf{x} = c$. Demuestra que $f(\ell)$ es la recta dada por la ecuación $f(\mathbf{n}) \cdot \mathbf{x} = c$ (ojo: tienes que demostrar dos contenciones). Concluye que f manda al haz paralelo con dirección $\mathbf{d}$ en el haz paralelo con dirección $f(\mathbf{d})$.

Un ejemplo

Para fijar ideas, concluyamos esta sección con un ejemplo que nos ayude a comprender el poder de los resultados obtenidos. Dibujo

Sea $f : \mathbb{R}^2 \to \mathbb{R}^2$ la rotación de 45° (o $\pi/4$) alrededor del punto $\mathbf{p} = (3, 2)$; el problema es escribir una fórmula explícita para f. Es decir, supongamos que queremos dar la función a una computadora para que haga algo con ella, que sepa aplicarla a puntos concretos; no le podemos dar la descripción que acabamos de hacer (y que cualquier ser humano entiende) para definir f, sino que tiene que ser mucho más concreta: una fórmula analítica.

Por la Proposición 3.10, sabemos que f se escribe como

$$f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{b},$$

para alguna $g \in \mathbf{O}(2)$ y alguna $\mathbf{b} \in \mathbb{R}^2$. El problema se divide entonces en dos: cómo se escribe g y quién es $\mathbf{b}$. Primero, g debe ser la rotación de $\pi/4$ en el origen pues f cambia las líneas horizontales en líneas a 45° y en la expresión anterior sólo g puede hacer esta gracia (pues una vez que apliquemos g la traslación mantendrá la orientación de las líneas). Por la Proposición 3.11, para expresar g basta ver que le hace a la base canónica $\mathbf{e}_1$, $\mathbf{e}_2$ y sabemos que tiene que ir a $\mathbf{u}$, $\mathbf{u}^\perp$, donde $\mathbf{u}$ es el vector unitario a 45°. Es decir, sea

$$\mathbf{u} = (\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}),$$

y entonces por (3.5) g se escribe

$$\begin{aligned} g(x,y) &= x\mathbf{u} + y\mathbf{u}^\perp \\ &= x(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}) + y(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}) \\ &= \frac{\sqrt{2}}{2}(x-y, x+y) = \frac{1}{\sqrt{2}}(x-y, x+y), \end{aligned}$$

y esto sí que lo entiende una computadora. Para convencernos de que vamos por buen camino basta checar que efectivamente esta fórmula da $g(\mathbf{e}_1) = \mathbf{u}$ y $g(\mathbf{e}_2) = \mathbf{u}^\perp$; e inclusive que $g(\mathbf{u}) = \mathbf{e}_2$ como dice la geometría.

El segundo problema es encontrar $\mathbf{b}$. Sabemos por la demostración de la Proposición 3.10 que $\mathbf{b} = f(\mathbf{0})$. Y entonces bastará encontrar este valor particular. Esto se puede razonar de varias maneras, pero lo haremos de manera muy general. Recordemos que definimos rotar en el punto $\mathbf{c}$ como trasladar $\mathbf{c}$ al origen, rotar ahí y luego regresar $\mathbf{c}$ a su lugar. En términos de funciones esto se escribe

$$f = \tau_{\mathbf{c}} \circ g \circ \tau_{-\mathbf{c}},$$

donde $\tau_{\mathbf{x}}$ es la traslación por $\mathbf{x}$. De aquí se podría sacar directamente la fórmula para f pues ya sabemos expresar g y las traslaciones, y falta nada más desarrollar las expresiones; o bien, podemos sólo aplicar esto al $\mathbf{0}$:

$$\begin{aligned} \mathbf{b} &= f(\mathbf{0}) = (\tau_{\mathbf{c}} \circ g \circ \tau_{-\mathbf{c}})(\mathbf{0}) \\ &= (\tau_{\mathbf{c}} \circ g)(\mathbf{0} - \mathbf{c}) = \tau_{\mathbf{c}}(g(-\mathbf{c})) = g(-\mathbf{c}) + \mathbf{c} \\ &= g(-3,-2) + (3,2) = \frac{1}{\sqrt{2}}(-1,-5) + (3,2) \\ &= (3 - \frac{1}{\sqrt{2}}, 2 - \frac{5}{\sqrt{2}}). \end{aligned}$$

De ahí concluimos que la expresión analítica de f es

$$\begin{aligned} f(x,y) &= g(x,y) + f(0) \\ &= \frac{1}{\sqrt{2}}(x-y, x+y) + (3 - \frac{1}{\sqrt{2}}, 2 - 5\frac{1}{\sqrt{2}}) \\ &= \frac{1}{\sqrt{2}}\left(x - y - 1 + 3\sqrt{2}\,,\ x + y - 5 + 2\sqrt{2}\right). \end{aligned}$$

EJERCICIO 3.45 Encuentra la expresión analítica de las siguientes isometrías:
a) La reflexión en la recta $\ell : x + y = 1$.
b) La rotación de $\pi/2$ en el punto $(-2, 3)$.
c) La rotación de π en el punto $(4, 3)$.
d) La reflexión en la recta $\ell : y = 1$ seguida de la traslación por $(2, 0)$.

EJERCICIO 3.46 Describe geométricamente (con palabras) las siguientes isometrías:
a) $f(x, y) = (x + 1, -y)$.
b) $f(x, y) = (-x + 2, -y)$.
c) $f(x, y) = (-y, -x + 2)$.

3.4 Las funciones lineales

Hemos visto que las transformaciones ortogonales de $\mathbb{R}^2$ quedan determinadas por lo que le hacen a la base canónica; es decir, que se escriben

$$f(x, y) = x\,\mathbf{u} + y\,\mathbf{v},$$

donde $\mathbf{u}$ y $\mathbf{v}$ son vectores fijos ($\mathbf{u} = f(\mathbf{e}_1)$ y $\mathbf{v} = f(\mathbf{e}_2)$, bien por definición o bien por la fórmula); y además, vimos que si la función es ortogonal entonces $\mathbf{u}$ y $\mathbf{v}$ forman una base ortonormal. Pero si en esta fórmula no pedimos nada a $\mathbf{u}$ y a $\mathbf{v}$, simplemente que sean vectores cualesquiera en $\mathbb{R}^2$, obtenemos una familia de funciones mucho más grande: las *lineales* de $\mathbb{R}^2$ en $\mathbb{R}^2$ (ver el siguiente teorema). Estas funciones son la base del álgebra lineal y serán el fundamento para las transformaciones que más nos interesan, así que vale la pena definirlas y estudiarlas un rato en general.

Definición 3.4.1 Una función $f : \mathbb{R}^n \to \mathbb{R}^m$ es *lineal* si para todos los vectores $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$ y todo número $t \in \mathbb{R}$ se cumple que:

$$\begin{aligned} &\textbf{i)} \quad f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}) \\ &\textbf{ii)} \quad f(t\,\mathbf{x}) = t\,f(\mathbf{x}). \end{aligned}$$

Estas funciones, al preservar la suma y la multiplicación por escalares, conservan las operaciones básicas de un espacio vectorial y por eso son tan importantes. Pero en el caso que nos ocupa, tenemos el siguiente teorema.

Teorema 3.13 *Una función* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *es lineal, si y sólo si se escribe*

$$f(x, y) = x\,\mathbf{u} + y\,\mathbf{v},$$

donde $\mathbf{u}$ *y* $\mathbf{v}$ *son vectores fijos en* $\mathbb{R}^2$.

Demostración. Supongamos que $f : \mathbb{R}^2 \to \mathbb{R}^2$ es lineal y sean $\mathbf{u} := f(\mathbf{e}_1)$ y $\mathbf{v} := f(\mathbf{e}_2)$, donde, recuérdese, $\mathbf{e}_1 = (1, 0)$ y $\mathbf{e}_2 = (0, 1)$ son la base canónica de $\mathbb{R}^2$. Dado cualquier $\mathbf{x} \in \mathbb{R}^2$, tenemos que si $\mathbf{x} = (x, y)$ entonces $\mathbf{x} = x\,\mathbf{e}_1 + y\,\mathbf{e}_2$, de donde, usando las propiedades (**i**) y (**ii**) respectivamente, se obtiene

$$\begin{aligned} f\,(x, y) &= f(\mathbf{x}) = f(x\,\mathbf{e}_1 + y\,\mathbf{e}_2) \\ &= f(x\,\mathbf{e}_1) + f(y\,\mathbf{e}_2) \\ &= x\,f(\mathbf{e}_1) + y\,f(\mathbf{e}_2) \\ &= x\,\mathbf{u} + y\,\mathbf{v}. \end{aligned}$$

Lo cual demuestra que cualquier función lineal se expresa en la forma deseada.

Por el otro lado, veamos que la función $f(x, y) = x\,\mathbf{u} + y\,\mathbf{v}$, con $\mathbf{u}$ y $\mathbf{v}$ arbitrarios, es lineal. Cambiando la notación de las coordenadas, sean $\mathbf{x} = (x_1, x_2)$ y $\mathbf{y} = (y_1, y_2)$. De la definición de f y las propiedades elementales de la suma vectorial tenemos

$$\begin{aligned} f\,(\mathbf{x} + \mathbf{y}) &= f\,(x_1 + y_1, x_2 + y_2) \\ &= (x_1 + y_1)\,\mathbf{u} + (x_2 + y_2)\,\mathbf{v} \\ &= (x_1\,\mathbf{u} + x_2\,\mathbf{v}) + (y_1\,\mathbf{u} + y_2\,\mathbf{v}) \\ &= f\,(\mathbf{x}) + f(\mathbf{y})\,, \end{aligned}$$

lo cual demuestra que f cumple (**i**). Análogamente, f cumple (**ii**), pues para cualquier $t \in \mathbb{R}$

$$\begin{aligned} f(t\,\mathbf{x}) &= (tx_1)\,\mathbf{u} + (tx_2)\,\mathbf{v} \\ &= t\,(x_1\,\mathbf{u} + x_2\,\mathbf{v}) = t(\,f(\mathbf{x})). \end{aligned}$$

Lo cual concluye la demostración del teorema. □

Animacion Interactiva

Tenemos entonces que las transformaciones ortogonales son funciones lineales, así que ya hemos visto varios ejemplos. Pero son muchas más las lineales de $\mathbb{R}^2$ en $\mathbb{R}^2$ que las ortogonales (véase el Ejercicio 3.50). Si tomamos la *cuadrícula canónica* con verticales y horizontales en los enteros de los ejes, una transformación lineal la manda en la "rombícula" que generan los vectores imágenes de la base canónica y que consiste en tomar las paralelas a uno de ellos que pasan por los múltiplos enteros del otro. Sólo cuando esta rombícula se sigue viendo como cuadrícula, y del mismo tamaño, tenemos una transformación ortogonal. Pero además, nótese que hemos definido funciones lineales entre diferentes espacios vectoriales y que en la demostración del

teorema no tuvimos que explicitar en dónde viven $\mathbf{u}$ y $\mathbf{v}$. Si vivieran en $\mathbb{R}^3$, por ejemplo, la imagen sería la misma, un plano rombiculado, que tiene la libertad extra de bambolearse.

EJERCICIO 3.47 Demuestra que una función $f : \mathbb{R}^n \to \mathbb{R}^m$ lineal preserva el origen, es decir, que cumple $f(\mathbf{0}) = \mathbf{0}$.

EJERCICIO 3.48 Demuestra que una función lineal manda líneas en líneas (de ahí su nombre) o a veces en puntos.

EJERCICIO 3.49 Demuestra que la composición de funciones lineales es lineal.

EJERCICIO 3.50 Demuestra que la función $f : \mathbb{R}^2 \to \mathbb{R}^2$, definida por $f(x, y) = x\,\mathbf{u} + y\,\mathbf{v}$ es ortogonal si y sólo si $\mathbf{u}$, $\mathbf{v}$ forman una base ortonormal.

EJERCICIO 3.51 Demuestra que la función $f : \mathbb{R} \to \mathbb{R}$ definida como $f(x) = 3x$ es una función lineal.

EJERCICIO 3.52 Demuestra que una función $f : \mathbb{R} \to \mathbb{R}$ es lineal si y sólo si es la función $f(x) = a\,x$ para alguna $a \in \mathbb{R}$.

EJERCICIO 3.53 Demuestra que la función $f : \mathbb{R} \to \mathbb{R}^2$ definida como $f(x) = (3x, 5x)$ es una función lineal.

EJERCICIO 3.54 Demuestra que una función $f : \mathbb{R} \to \mathbb{R}^2$ es lineal si y sólo si se escribe $f(x) = (ax, bx)$, para algunos a, $b \in \mathbb{R}$.

EJERCICIO 3.55 Demuestra que la función $f : \mathbb{R}^2 \to \mathbb{R}$ definida como $f(x, y) = 2x + 4y$ es una función lineal.

EJERCICIO 3.56 Demuestra que la función $f : \mathbb{R}^2 \to \mathbb{R}^2$ definida como $f(x, y) = (x + y, y - x)$ es una función lineal.

EJERCICIO 3.57 Exhibe una función lineal $f : \mathbb{R}^2 \to \mathbb{R}^2$ que mande el eje x a la recta $\ell : 2x - y = 0$.

EJERCICIO 3.58 Demuestra que una función $f : \mathbb{R}^3 \to \mathbb{R}^n$ es lineal, si y sólo si se escribe

$$f(x, y, z) = x\,\mathbf{u} + y\,\mathbf{v} + z\,\mathbf{w},$$

donde $\mathbf{u} = f(\mathbf{e}_1)$, $\mathbf{v} = f(\mathbf{e}_2)$ y $\mathbf{w} = f(\mathbf{e}_3)$.

3.4.1 Extensión lineal

De los ejercicios anteriores, en particular el último, debe quedar claro que el Teorema 3.13 es sólo un caso particular de un resultado general que haremos explícito en esta sección. Aunque estemos especialmente interesados en las dimensiones $1, 2, 3$, para ocuparnos de todas ellas a la vez, y evitarnos el molesto trabajo de 6, seis, casos particulares, es necesario trabajar con los abstractos $\mathbb{R}^n$ y $\mathbb{R}^m$.

Como definimos en el capítulo 1, dados los vectores $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ en $\mathbb{R}^n$, una

combinación lineal de ellos es el vector

$$\sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i \in \mathbb{R}^n,$$

donde $\lambda_i \in \mathbb{R}$, $i = 1, ..., k$, son escalares cualesquiera llamados los *coeficientes* de la combinación lineal. Por ejemplo, todas las combinaciones lineales de un solo vector ($k = 1$) forman la recta que genera (si es no nulo) y las de dos vectores forman un plano (si no son paralelos). En una combinación lineal se están utilizando simultáneamente las dos operaciones básicas de un espacio vectorial. Ahora bien, si tenemos una función lineal $f : \mathbb{R}^n \to \mathbb{R}^m$, se cumple que *preserva combinaciones lineales* pues

$$f(\sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i) = \sum_{i=1}^{k} f(\lambda_i \boldsymbol{v}_i) = \sum_{i=1}^{k} \lambda_i f(\boldsymbol{v}_i),$$

donde se usa la propiedad **(i)** de las funciones lineales en la primera igualdad y la **(ii)** en la segunda. Así que podemos redefinir función lineal como aquella que cumple

$$f(\sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i) = \sum_{i=1}^{k} \lambda_i f(\boldsymbol{v}_i), \tag{3.6}$$

para cualesquiera $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k \in \mathbb{R}^n$ y $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{R}$; pues esta condición claramente incluye la **(i)** y la **(ii)** de la definición original. Hay que hacer énfasis en que en esta igualdad la combinación lineal de la derecha se da en $\mathbb{R}^m$ (el codominio), mientras que la que tenemos en el lado izquierdo ocurre en $\mathbb{R}^n$ (el dominio), antes de aplicarle la función, que no necesariamente son el mismo espacio.

Por otro lado, $\mathbb{R}^n$ tiene una *base canónica*: los vectores $\boldsymbol{e}_1 = (1, 0, ..., 0)$, $\boldsymbol{e}_2 = (0, 1, ..., 0), ..., \boldsymbol{e}_n = (0, 0, ..., 1)$; es decir, para $i = 1, ..., n$ el vector $\boldsymbol{e}_i \in \mathbb{R}^n$ tiene i-ésima coordenada 1 y el resto 0. Estos se llaman canónicos pues la combinación lineal de ellos que nos da cualquier otro vector $\boldsymbol{x} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ se obtiene trivialmente ("canónicamente") de sus coordenadas:

$$(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} x_i \, \boldsymbol{e}_i.$$

De tal manera que para cualquier función lineal f que sale de $\mathbb{R}^n$ se tiene, usando (3.6) y esta última expresión, que

$$f(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} x_i \, f(\boldsymbol{e}_i).$$

Nos falta ver que los valores que puede tomar f en la base canónica son arbitrarios. Es decir, si tomamos n vectores $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n \in \mathbb{R}^m$ arbitrariamente, y queremos

encontrar una función lineal $f : \mathbb{R}^n \to \mathbb{R}^m$ que cumpla $f(\mathbf{e}_i) = \mathbf{u}_i$ para $i = 1, 2, \ldots, n$, entonces debemos definirla en todo $\mathbb{R}^n$ como

$$f(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} x_i \mathbf{u}_i.$$

A esta función se le llama la *extensión lineal* a $\mathbb{R}^n$ de lo que determinamos para la base canónica. Pero nos falta demostrar que efectivamente es lineal. Sin embargo, esto es muy fácil de hacer usando la definición original de función lineal y las de suma y multiplicación por escalares; hay que copiar la del Teorema 3.13 con más coordenadas, y dejamos este trabajo al lector. Podemos resumir entonces con el siguiente teorema.

Teorema 3.14 *Una función* $f : \mathbb{R}^n \to \mathbb{R}^m$ *es lineal si y sólo si se escribe*

$$f(x_1, x_2, ..., x_n) = x_1 \mathbf{u}_1 + x_2 \mathbf{u}_2 + \cdots + x_n \mathbf{u}_n = \sum_{i=1}^{n} x_i \mathbf{u}_i,$$

donde $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n \in \mathbb{R}^m$. *Nótese que* $\mathbf{u}_i = f(\mathbf{e}_i)$. □

Corolario 3.15 *Si* $f : \mathbb{R}^n \to \mathbb{R}^m$ *y* $g : \mathbb{R}^n \to \mathbb{R}^m$ *son dos funciones lineales tales que* $f(\mathbf{e}_i) = g(\mathbf{e}_i)$ *para* $i = 1, 2, \ldots, n$, *entonces* $f(\mathbf{x}) = g(\mathbf{x})$ *para toda* $\mathbf{x} \in \mathbb{R}^n$.

Demostración. Por el teorema, tienen la misma expresión. □

El teorema, o bien su corolario, pueden llamarse "Teorema de extensión única de funciones lineales". La moraleja es que una función lineal depende únicamente de unos cuantos valores: los que le asigna a la base canónica, y éstos, una vez fijado el codominio, son arbitrarios.

EJERCICIO 3.59 Describe el lugar geométrico definido como $\mathcal{L} = \{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}\}$, donde $f : \mathbb{R} \to \mathbb{R}^3$ es una función lineal.
EJERCICIO 3.60 Describe el lugar geométrico definido como $\Pi = \{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2\}$, donde $f : \mathbb{R}^2 \to \mathbb{R}^3$ es una función lineal.
EJERCICIO 3.61 ¿Qué ejercicios del bloque anterior se siguen inmediatamente del Teorema 3.14? Reescríbelos.

3.4.2 La estructura de las funciones lineales

En el capítulo 1 dejamos en suspenso un ejercicio donde se pregunta al lector si conoce algún otro espacio vectorial que no sea $\mathbb{R}^n$. Sería desleal dejar pasar ese ejemplo cuando lo tenemos en las narices. Efectivamente, cuando fijamos el dominio y el contradominio, las funciones lineales entre ellos tienen naturalmente la estructura de un espacio vectorial.

Fijemos la notación. Sea $\mathcal{L}(n,m)$ el conjunto de todas las funciones lineales de $\mathbb{R}^n$ a $\mathbb{R}^m$, es decir

$$\mathcal{L}(n,m) = \{f : \mathbb{R}^n \to \mathbb{R}^m \mid f \text{ es lineal}\}.$$

Como en el codominio se tienen las operaciones de suma y multiplicación por escalares, estas operaciones se pueden definir también en las funciones; en cierta manera, las funciones las heredan. Dadas $f, g \in \mathcal{L}(n,m)$ y $t \in \mathbb{R}$, sean

$$\begin{gathered} f + g : \mathbb{R}^n \to \mathbb{R}^m \\ (f+g)(\mathbf{x}) := f(\mathbf{x}) + g(\mathbf{x}) \end{gathered}$$

y

$$\begin{gathered} t\,f : \mathbb{R}^n \to \mathbb{R}^m \\ (t\,f)(\mathbf{x}) := t(f(\mathbf{x})). \end{gathered}$$

Hay que demostrar que estas nuevas funciones también son lineales. Y esto es muy fácil a partir de la definición original, pues dados $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ y $\lambda, \mu \in \mathbb{R}$,

$$\begin{aligned} (f+g)(\lambda\mathbf{x} + \mu\mathbf{y}) &= f(\lambda\mathbf{x} + \mu\mathbf{y}) + g(\lambda\mathbf{x} + \mu\mathbf{y}) \\ &= \lambda f(\mathbf{x}) + \mu f(\mathbf{y}) + \lambda g(\mathbf{x}) + \mu g(\mathbf{y}) \\ &= \lambda(f(\mathbf{x}) + g(\mathbf{x})) + \mu(f(\mathbf{y}) + g(\mathbf{y})) \\ &= \lambda(f+g)(\mathbf{x}) + \mu(f+g)(\mathbf{y}). \end{aligned}$$

Formalmente nos falta demostrar que estas operaciones cumplen todas las propiedades que se les exigen a los espacios vectoriales (el Teorema **??**). Esto depende esencialmente de que $\mathbb{R}^m$ es espacio vectorial (no de que sean funciones lineales) y dejamos los detalles como ejercicio lateral (no esencial) a la línea del texto. Los nuevos ejemplos de espacio vectorial son entonces los conjuntos de funciones que salen de algún lugar (con alguna propiedad —el caso que desarrollamos fue que son lineales—) pero que **caen** en un espacio vectorial (en nuestro caso, $\mathbb{R}^m$).

EJERCICIO 3.62 Sea V un espacio vectorial y sea X cualquier conjunto. Demuestra que el conjunto de todas las funciones de X en V, $\mathcal{F}(X,V)$, tiene una estructura natural de espacio vectorial. (En particular $\mathcal{L}(n,m)$ es un subespacio vectorial de $\mathcal{F}(\mathbb{R}^n, \mathbb{R}^m)$ que es inmensamente más grande.)

EJERCICIO 3.63 Sea $\Delta_n = \{1, 2, ..., n\}$ el conjunto finito "canónico" con n elementos. Como corolario del Teorema 3.14, demuestra que los espacios vectoriales $\mathcal{L}(n,m)$ y $\mathcal{F}(\Delta_n, \mathbb{R}^m)$ se pueden identificar naturalmente.

3.5 Matrices

Las matrices son arreglos rectangulares de números, tablas podría decirse, donde los renglones y las columnas no tienen ningún significado extra, como podrían tener, por ejemplo, en la tabla de calificaciones parciales de alumnos en un curso. Así de simple, las matrices son tablas limpias, abstractas. En nuestro contexto actual, habrán de ser los "paquetes" que cargan toda la información de las funciones lineales y nos darán la herramienta para hacer cálculos respecto a ellas, además de simplificar la notación.

En matemáticas, la notación es importante. Una notación clara y sencilla permite ver la esencia de las cosas, de las ideas y de los problemas. Por el contrario, con una notación complicada o confusa es fácil perderse en la labor de desentrañarla y nunca llegar a las ideas profundas; piénsese, por ejemplo, en diseñar un algoritmo para multiplicar con números romanos. Viene esto al caso, pues esta sección trata básicamente de cómo simplificar la notación para lograr manejar las transformaciones geométricas muy en concreto. Encontraremos la mecánica y la técnica para hacer cálculos con facilidad. Para empezar, cambiaremos nuestra notación de vectores.

3.5.1 Vectores columna

Como crítica a la notación que traemos, escribamos una función lineal $f : \mathbb{R}^3 \to \mathbb{R}^3$ explícitamente. Digamos que $f(\boldsymbol{e}_1) = (1,2,3)$, $f(\boldsymbol{e}_2) = (6,5,4)$ y $f(\boldsymbol{e}_3) = (2,3,1)$; entonces sabemos que f se escribe

$$\begin{aligned} f(x,y,z) &= x(1,2,3) + y(6,5,4) + z(2,3,1) \\ &= (x + 6y + 2z, 2x + 5y + z, 3x + 4y + z). \end{aligned}$$

Para hacer este cálculo, la vista tuvo que andar brincoteando en una línea, buscando y contando comas; es muy fácil equivocarse. Es más, ¿dónde está el error? Sin embargo, sabemos que el cálculo es sencillísimo, mecánico. Si en lugar de considerar los vectores como renglones los tomamos como columnas, el mismo cálculo se hace evidente a la vista

$$f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + y\begin{pmatrix} 6 \\ 5 \\ 4 \end{pmatrix} + z\begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} x + 6y + 2z \\ 2x + 5y + 3z \\ 3x + 4y + z \end{pmatrix} \tag{3.7}$$

y eliminamos las comas (y corregimos el error). Ahora cada renglón se refiere a una coordenada, y se ve todo de golpe; mucho mejor notación que haremos oficial.

Notación 3.1 De ahora en adelante, los vectores en $\mathbb{R}^n$ se escribirán como columnas y no como renglones. Es decir, si antes escribíamos $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, ahora escribiremos

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

El inconveniente de que entonces se complica escribir un vector dentro del texto (aquí, por ejemplo), lo subsanamos al llamar a los vectores renglón, *transpuestos* de los vectores columna, y los denotaremos con $\mathbf{x}^\top = (x_1, x_2, \ldots, x_n)$; o bien, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$ conviniendo que los *transpuestos* de vectores renglón son los vectores. Aunque a veces se nos va a olvidar poner el exponente $\top$ de "transpuesto" y el lector nos lo perdonará; y con el tiempo nos lo agradecerá.

Así, por ejemplo, tenemos que la base canónica de $\mathbb{R}^2$ es $\mathbf{e}_1 = (1, 0)^\top$ y $\mathbf{e}_2 = (0, 1)^\top$, que quiere decir

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{y} \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Vale la pena insistir en que no hay ningún cambio esencial. Sólo cambiamos de convención, una pareja ordenada de números se puede pensar horizontal (de izquierda a derecha) o vertical (de arriba a abajo); estamos conviniendo en que lo haremos y escribiremos en sentido vertical.

3.5.2 La matriz de una función lineal

Una matriz de $m \times n$ es un arreglo rectangular (o tabla) de números reales con m renglones y n columnas. Si, usando dos subíndices, denotamos con a_{ij} al número que está en el renglon i y la columna j, llamado *entrada*, tenemos que una matriz de $m \times n$ es

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Podríamos también definirla como un conjunto ordenado de n vectores en $\mathbb{R}^m$: sus columnas. Es decir, la matriz A anterior se puede escribir como

$$A = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n), \quad \text{donde} \quad \mathbf{u}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix} \in \mathbb{R}^m, \; i = 1, 2, \ldots, n.$$

De tal manera que, de acuerdo con el Teorema 3.14, una matriz $m \times n$ tiene justo la información de una función lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$ (ojo: se invierte el orden por la convención de que las funciones se componen "hacia atras"). Explícitamente, a la matriz A se le asocia la función lineal $f : \mathbb{R}^n \to \mathbb{R}^m$ que manda al vector canónico $\mathbf{e}_i \in \mathbb{R}^n$ en su columna i-ésima, es decir, tal que $f(\mathbf{e}_i) = \mathbf{u}_i$ para $i = 1, 2, \ldots, n$. O

bien, a la inversa, a cada función lineal $f : \mathbb{R}^n \to \mathbb{R}^m$ se le asocia la matriz $m \times n$ que tiene como columnas a sus valores en la base canónica, es decir, se le asocia la matriz

$$A = (f(\mathbf{e}_1), f(\mathbf{e}_2), \ldots, f(\mathbf{e}_n)) .$$

Por ejemplo, a la función lineal de $\mathbb{R}^3$ en $\mathbb{R}^3$ definida en (3.7) se le asocia la matriz

$$\begin{pmatrix} 1 & 6 & 2 \\ 2 & 5 & 3 \\ 3 & 4 & 1 \end{pmatrix}$$

que es ya una compactación considerable en la notación; a la transformación "compadre ortogonal" de $\mathbb{R}^2$ en $\mathbb{R}^2$ se le asocia la matriz

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

y a la función lineal que a cada vector en $\mathbb{R}^3$ lo manda a su segunda coordenada (en $\mathbb{R}$) se le asocia la matriz 1×3: $\begin{pmatrix} 0 & 1 & 0 \end{pmatrix} = (0, 1, 0)$.

Ya logramos nuestro primer objetivo, empaquetar en una matriz toda la información de una función lineal. Ahora usémosla como herramienta. Primero, definiremos el producto de una matriz por un vector para que el resultado sea lo que su función lineal asociada hace al vector. Es decir, si A es una matriz $m \times n$, podrá multiplicar sólo a vectores $\mathbf{x} \in \mathbb{R}^n$ y el resultado será un vector en $\mathbb{R}^m$. Como ya vimos, A se puede escribir $A = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ donde $\mathbf{u}_i \in \mathbb{R}^m$, para $i = 1, 2, \ldots, n$; y por su parte, sea $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\top$. Definimos entonces el *producto* de A por $\mathbf{x}$ como

$$A\mathbf{x} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \cdots + x_n\mathbf{u}_n, \tag{3.8}$$

de tal manera que el Teorema 3.14 junto con su corolario y lo que hemos visto de matrices se puede resumir en el siguiente teorema.

Teorema 3.16 *Las funciones lineales de* $\mathbb{R}^n$ *en* $\mathbb{R}^m$ *están en correspondencia natural y biunívoca con las matrices de* $m \times n$*, de tal manera que a la función* f *le corresponde la matriz* A *que cumple*

$$f(\mathbf{x}) = A\mathbf{x}$$

para todo $\mathbf{x} \in \mathbb{R}^n$*.*

Obsérvese que cuando $m = 1$ el producto que acabamos de definir corresponde a nuestro viejo conocido, el producto interior. Es decir, si $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ entonces

$$\mathbf{y}^\top \mathbf{x} = \mathbf{y} \cdot \mathbf{x}.$$

De tal manera que si escribimos A como una columna de renglones, en vez de un renglón de columnas, entonces hay vectores $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$ (los renglones) tales que

$$Ax = \begin{pmatrix} v_1^\top \\ v_2^\top \\ \vdots \\ v_m^\top \end{pmatrix} x = \begin{pmatrix} v_1 \cdot x \\ v_2 \cdot x \\ \vdots \\ v_m \cdot x \end{pmatrix}.$$

Esta definición del producto de una matriz por un vector es, aunque equivalente, más "desagradable" que la anterior, pues los vectores $v_i \in \mathbb{R}^n$ no tienen ningún significado geométrico; son simplemente la colección ordenada de las i-ésimas coordenadas de los vectores $u_j \in \mathbb{R}^m$. Aunque somos injustos, sí tienen un significado, los renglones $v_i^\top$ son las matrices asociadas a las m funciones lineales $\mathbb{R}^n \to \mathbb{R}$ que dan las coordenadas de la función original.

Para fijar ideas, y no perdernos en las abstracciones, terminamos esta sección con la expresión explícita del caso que más nos interesa en este libro: la multiplicación de una matriz de 2×2 por un vector en $\mathbb{R}^2$, que tiene la fórmula general

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$$

donde todos los protagonistas son simples números.

EJERCICIO 3.64 Encuentra las matrices asociadas a las funciones lineales de algunos de los ejercicios anteriores.

EJERCICIO 3.65 Un vector $v \in \mathbb{R}^n$ se puede pensar como matriz $n \times 1$. Como tal, ¿a qué función lineal representa según el Teorema 3.16?

EJERCICIO 3.66 El conjunto de todas las funciones lineales de $\mathbb{R}^n$ en $\mathbb{R}^m$, que denotamos con $\mathcal{L}(n, m)$ en la Sección 3.4.2, tienen la estructura de espacio vectorial y el Teorema 3.16 dice que está en biyección natural con las matrices $m \times n$. Si definimos la suma de matrices y la multiplicación de escalares por matrices de modo que correspondan a las operaciones análogas en $\mathcal{L}(n, m)$, demuestra que $A + B$ (que tiene sentido sólo cuando A y B son del mismo tipo $m \times n$) es sumar entrada por entrada, y que tA es multiplicar todas las entradas por t. ¿Puedes dar una demostración fácil de que $\mathcal{L}(n, m)$ es un espacio vectorial? (Observa, del ejercicio anterior, que $\mathcal{L}(1, n)$ se identifica naturalmente con $\mathbb{R}^n$.)

3.5.3 Multiplicación de matrices

Vamos ahora a definir la multiplicación de matrices correspondiendo a la composición de funciones lineales. Tiene sentido componer dos funciones sólo cuando una "acaba" donde la otra "empieza"; más formalmente, cuando el codominio de una coincide con el dominio de la otra. De igual forma, sólo tendrá sentido multiplicar las matrices
Dibujo cuyas dimensiones se "acoplen". Veámoslo.

Sean $f : \mathbb{R}^n \to \mathbb{R}^m$ y $g : \mathbb{R}^m \to \mathbb{R}^k$ dos funciones lineales. Por el Ejercicio 3.49, $g \circ f : \mathbb{R}^n \to \mathbb{R}^k$ también es lineal. Sean A la matriz $m \times n$ y B la matriz $k \times m$ que corresponden a f y a g respectivamente. *Definimos el producto* BA como la matriz $k \times n$ que corresponde a la función lineal $g \circ f$. Es decir, BA es la única matriz $k \times n$ que cumple

$$(g \circ f)(\mathbf{x}) = (BA)\mathbf{x} \quad \text{para todo} \quad \mathbf{x} \in \mathbb{R}^n.$$

Aunque esta definición ya es precisa según el Teorema 3.16, todavía no nos dice cómo multiplicar. Esto habrá que deducirlo. Recordemos que las columnas de una matriz son las imágenes de la base canónica bajo la función asociada. Así que si $A = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ donde $\mathbf{u}_i = f(\mathbf{e}_i) \in \mathbb{R}^m$, entonces $(g \circ f)(\mathbf{e}_i) = g(f(\mathbf{e}_i)) = g(\mathbf{u}_i) = B\mathbf{u}_i$. Y por lo tanto

$$BA = B\,(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = (B\mathbf{u}_1, B\mathbf{u}_2, \ldots, B\mathbf{u}_n)\,.$$

Lo cual es muy natural: para obtener las columnas de la nueva matriz, se usa la multiplicación de B por los vectores columna de A, multiplicación que ya definimos. Para expresar cada una de las entradas de la matriz BA, habrá que expresar B como una columna de vectores renglon, y se obtiene

$$BA = \begin{pmatrix} \mathbf{w}_1^\top \\ \mathbf{w}_2^\top \\ \vdots \\ \mathbf{w}_k^\top \end{pmatrix} (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{u}_1 & \mathbf{w}_1 \cdot \mathbf{u}_2 & \cdots & \mathbf{w}_1 \cdot \mathbf{u}_n \\ \mathbf{w}_2 \cdot \mathbf{u}_1 & \mathbf{w}_2 \cdot \mathbf{u}_2 & \cdots & \mathbf{w}_2 \cdot \mathbf{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{w}_k \cdot \mathbf{u}_1 & \mathbf{w}_k \cdot \mathbf{u}_2 & \cdots & \mathbf{w}_k \cdot \mathbf{u}_n \end{pmatrix}, \tag{3.9}$$

En esta fórmula queda claro por qué los renglones de B (los transpuestos de los vectores $\mathbf{w}_i$) y las columnas de A (los vectores $\mathbf{u}_j$) tienen que estar en el mismo espacio ($\mathbb{R}^m$); y cuál es la mecánica[1] para obtener las entradas de una matriz $k \times n$ (BA) a partir de una $k \times m$ (B) y una $m \times n$ (A). Esta fórmula da, en el caso de matrices 2×2, lo siguiente

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} a\alpha + b\gamma & a\beta + b\delta \\ c\alpha + d\gamma & c\beta + d\delta \end{pmatrix}.$$

EJERCICIO 3.67 Sean

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 3 \\ -2 & 2 \end{pmatrix}, \quad C = \begin{pmatrix} -2 & 0 \\ 1 & 2 \end{pmatrix}.$$

Calcula AB, $A(B + C)$, $(C - B)A$.

[1] La mecánica de la multiplicación de matrices corresponde a la *ley del karatazo*: un karateca en guardia pone su antebrazo derecho vertical y el izquierdo horizontal —¡$(huuouui)$!— y se apresta a multiplicar dos matrices: recoge con la mano derecha una columna de la matriz derecha y —¡$(aaah_1, aah_2, \cdots, ah_k)^\top$!— le va pegando karatazos a los renglones de la izquierda para obtener la columna correspondiente en el producto; y si es zurdo puede recoger con su buena renglones de la izquierda y —¡$(uuuf_1, uuf_2, \cdots, uf_n)$!— golpear a las columnas derechas (codo arriba) para ir sacando, de una en una, las entradas del producto.

EJERCICIO 3.68 Exhibe matrices A y B de 2×2 tales que $AB \neq BA$. De tal manera que el producto de matrices no es conmutativo.

EJERCICIO 3.69 Exhibe matrices A y B de 2×2 tales que $AB = 0$, pero $A \neq 0$ y $B \neq 0$; donde 0 es la matriz cero (con todas sus entradas 0). *(Piensa en términos de funciones.)*

EJERCICIO 3.70 Demuestra que si A, B, C son matrices 2×2, entonces $A(BC) = (AB)C$, es decir, que el producto de matrices es asociativo, y por lo tanto tiene sentido escribir ABC. *(No necesariamente tienes que escribir todo el producto y tal vez tu demostración valga en general.)*

EJERCICIO 3.71 Demuestra que si A, B, C son matrices 2×2, entonces $A(B + C) = AB + AC$ y $(A + B)C = AC + BC$.

EJERCICIO 3.72 Demuestra que si A, B son matrices 2×2 entonces $A(kB) = (kA)B = k(AB)$ para cualquier $k \in \mathbb{R}$.

EJERCICIO 3.73 Usando que la composición de funciones es asociativa demuestra que si A, B, C son matrices $n \times n$ entonces $A(BC) = (AB)C$.

3.5.4 Algunas familias distinguidas de matrices

La matriz identidad

Aunque propiamente dicho no sea una familia se merece su párrafo aparte y, además, ser el primero. La *matriz identidad,* o simplemente "la identidad", es la matriz asociada a la función identidad, que se denota por I, y es la matriz que tiene 1 en la diagonal y 0 fuera de ella, es decir

$$I := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Cumple además que $AI = A$ e $IA = A$ para cualquier matriz A que se deje multiplicar por ella: es la identidad multiplicativa. Nótese que en realidad hay una identidad para cada dimensión, empezando por el 1; así que deberíamos escribir algo así como I_n, pero el contexto en que se usa siempre trae la información de la dimensión.

Homotecias

Las homotecias, como funciones, son simples "cambios de escala". Si tomamos un número $k \in \mathbb{R}$, $k \neq 0$, la función $f : \mathbb{R}^n \to \mathbb{R}^n$ definida por $f(x) = kx$ es claramente lineal y se llama una *homotecia*. Entonces tiene una matriz asociada que es kI, la matriz que tiene sólo k en la diagonal y 0 fuera de ella. Como función, lo que hace es

expander uniformemente desde el origen (si $k > 1$), o bien contraer (cuando $k < 1$); son un *zoom*, pues.

Las homotecias tienen la propiedad de conmutar con cualquier otra matriz, pues $A(kI) = k(AI) = kA = k\,(IA) = (kI)A$ (y aquí, pudo haber sido que las dos I sean diferentes, lo que depende de A). En términos de funciones, el cambio de escala se puede hacer antes o después de una función lineal y da lo mismo. Y además son las únicas matrices (pensando en matrices cuadradas) que tienen esta propiedad; pero esto lo dejamos como ejercicio.

EJERCICIO 3.74 Demuestra que si una matriz A de 2×2 es tal que $AB = BA$ para todas las matrices B de 2×2, entonces $A = kI$ para alguna $k \in \mathbb{R}$. (*Tienes que encontrar matrices apropiadas* B *que te vayan dando información; quizás conviene ver el siguiente párrafo y regresar.*)

Matrices de permutaciones

Recordemos de la Sección 3.1 que una permutación de n elementos es una función biyectiva $\rho : \Delta_n \to \Delta_n$, donde podemos tomar $\Delta_n = \{1, 2, \ldots, n\}$; y que todas estas permutaciones forman el grupo simétrico de orden n, $\mathbf{S}_n$, con $n!$ elementos. Ahora bien, a cada permutación $\rho : \Delta_n \to \Delta_n$ se le puede asociar la función lineal $f_\rho : \mathbb{R}^n \to \mathbb{R}^n$ definida por $f_\rho(\boldsymbol{e}_i) = \boldsymbol{e}_{\rho(i)}$; es decir, f_ρ permuta a los elementos de la base canónica según ρ y se extiende linealmente a $\mathbb{R}^n$. A la matriz asociada a esta función lineal se le llama la matriz *de* la permutación ρ; denotémosla A_ρ. Es fácil ver que la multiplicación de matrices de permutaciones es de nuevo una matriz de permutación y corresponde a la composición de las permutaciónes correspondientes. Es decir, si $\rho, \sigma \in \mathbf{S}_n$, entonces $A_\rho A_\sigma = A_{\rho\circ\sigma}$. Se tiene entonces que el grupo simétrico se puede ver como un "grupo de matrices".

Otra manera equivalente de definirlas es que son las matrices cuadradas de ceros y unos tales que tienen exactamente un 1 en cada renglón y en cada columna. Por ejemplo, la matriz

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

corresponde a la permutación $\{1 \mapsto 3, 2 \mapsto 2, 3 \mapsto 4, 4 \mapsto 1\}$.

EJERCICIO 3.75 Escribe todas las matrices de permutaciones de orden 2.

EJERCICIO 3.76 Escribe todas las matrices de permutaciones de orden 3. Identifica sus inversas (es decir la matriz asociada a la permutación inversa), y describe en cada caso la transformación lineal de $\mathbb{R}^3$ correspondiente.

EJERCICIO 3.77 Sea A_ρ la matriz de la permutación $\rho : \Delta_n \to \Delta_n$, y sea B cualquier otra matriz de $n \times n$. Describe a las matrices BA_ρ y $A_\rho B$ en términos de la permutación ρ.

***EJERCICIO 3.78** Sea C el cubo en $\mathbb{R}^3$ centrado en el origen, de lado 2 y con lados paralelos a los ejes coordenados; es decir, sus vértices son los ocho puntos cuyas tres coordenadas son 1 o -1. Describe el conjunto de matrices cuyas funciones lineales asociadas mandan al cubo en sí mismo. En particular, ¿cuántas son? *(Quizás valga la pena empezar con el problema análogo en dimensión* 2.)

Rotaciones

Dibujo

Regresamos, después de un provechoso paréntesis, a la motivación original que nos llevó a adentrarnos en las funciones lineales.

Sabemos que una rotación es una transformación lineal que manda al vector canónico $\mathbf{e}_1$ en un vector unitario $\mathbf{u} \in \mathbb{S}^1$ y al otro vector canónico $\mathbf{e}_2$ en su compadre ortogonal $\mathbf{u}^\perp$. Si escribimos a $\mathbf{u}$ en términos de su ángulo respecto del eje x, es decir, respecto a $\mathbf{e}_1$, entonces obtenemos que la matriz asociada a la rotación por un ángulo θ es

$$R_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Primero veamos que la rotación de un ángulo $-\theta$ nos regresa a la identidad. Como

$$\cos(-\theta) = \cos\theta \quad \text{y} \quad \sin(-\theta) = -\sin\theta \tag{3.10}$$

entonces

$$\begin{aligned} R_{-\theta}R_\theta &= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\theta + \sin^2\theta & \cos\theta(-\sin\theta) + \sin\theta\cos\theta \\ -\sin\theta\cos\theta + \cos\theta\sin\theta & \sin^2\theta + \cos^2\theta \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I\,. \end{aligned}$$

Podemos obtener las fórmulas trigonométricas clásicas para el coseno y el seno de la suma de ángulos como consecuencia de la composición de funciones y la multiplicación de matrices. Está claro que si rotamos un ángulo β y después un ángulo α, habremos rotado un ángulo $\alpha + \beta$. Así que

$$R_\alpha R_\beta = R_{\alpha+\beta}.$$

Pero por otro lado, al multiplicar las matrices tenemos

$$\begin{aligned} R_\alpha R_\beta &= \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} = \\ &= \begin{pmatrix} \cos\alpha\cos\beta - \sin\alpha\sin\beta & -\cos\alpha\sin\beta - \sin\alpha\cos\beta) \\ \sin\alpha\cos\beta + \cos\alpha\sin\beta & -\sin\alpha\sin\beta + \cos\alpha\cos\beta \end{pmatrix}. \end{aligned}$$

Y por lo tanto

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \qquad (3.11)$$
$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta,$$

pues si dos matrices coinciden ($R_{\alpha+\beta}$ y $R_\alpha R_\beta$) lo hacen entrada por entrada. Junto con las igualdades trigonométricas de inversos de ángulos, estas últimas dan

$$\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta \qquad (3.12)$$
$$\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta.$$

EJERCICIO 3.79 Da una expresión numérica bonita de las matrices asociadas a las rotaciones de $\pi/6, \pi/4, \pi/3, 2\pi/3, \pi, -\pi/2$.

EJERCICIO 3.80 Demuestra que

$$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha$$
$$\sin 2\alpha = 2\sin\alpha\cos\alpha.$$

Reflexiones

Animacion

Conviene parametrizar las reflexiones no por el ángulo de la imagen de $\mathbf{e}_1$, sino por su mitad, que es el ángulo de la recta-espejo de la reflexión. Pues si, como es natural, a una recta por el origen le asociamos su ángulo (entre 0 y π) con el rayo positivo del eje x, entonces la reflexión, E_θ, en la recta con ángulo θ manda a $\mathbf{e}_1$ en el vector unitario de ángulo 2θ, y por lo tanto su matriz asociada es

$$E_\theta := \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix},$$

pues el segundo vector no es el compadre ortogonal sino su negativo. Es muy fácil ver que $E_\theta E_\theta = I$, que corresponde a que una reflexión es su propia inversa.

Podemos ahora ver cómo se comporta la composición de reflexiones:

$$\begin{aligned} & E_\alpha E_\beta = \\ = & \begin{pmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{pmatrix} \begin{pmatrix} \cos 2\beta & \sin 2\beta \\ \sin 2\beta & -\cos 2\beta \end{pmatrix} \\ = & \begin{pmatrix} \cos 2\alpha\cos 2\beta + \sin 2\alpha\sin 2\beta & \cos 2\alpha\sin 2\beta - \sin 2\alpha\cos 2\beta \\ \sin 2\alpha\cos 2\beta - \cos 2\alpha\sin 2\beta & \sin 2\alpha\sin 2\beta + \cos 2\alpha\cos 2\beta \end{pmatrix} \\ = & \begin{pmatrix} \cos 2(\alpha-\beta) & -\sin 2(\alpha-\beta) \\ \sin 2(\alpha-\beta) & \cos 2(\alpha-\beta) \end{pmatrix} = R_{2(\alpha-\beta)} \end{aligned}$$

donde hemos usado (3.12). Así que la composición de dos reflexiones es la rotación del doble del ángulo entre sus espejos. Tal y como lo habíamos visto al hablar de generadores de grupos diédricos.

EJERCICIO 3.81 Da una expresión numérica bonita de las matrices asociadas a las reflexiones en espejos a $0, \pi/6, \pi/4, \pi/3, \pi/2, 2\pi/3, 3\pi/4$.

EJERCICIO 3.82 ¿Quiénes son $E_\alpha R_\beta$ y $R_\beta E_\alpha$?

EJERCICIO 3.83 Demuestra con matrices que $R_\beta E_0 R_{-\beta} = E_\beta$.

EJERCICIO 3.84 Demuestra que la fórmula (3.3) obtenida en la Sección 3.3.1 para la reflexión φ_ℓ en la línea ℓ, coincide con la que nos da una matriz E_θ cuando la línea ℓ pasa por el origen.

Matrices ortogonales (y transpuestas)

Decimos que una matriz es *ortogonal* si su función lineal asociada es ortogonal. Según nuestro estudio de las transformaciones ortogonales en $\mathbb{R}^2$, las matrices ortogonales de 2×2 son las que acabamos de describir: las rotaciones y las reflexiones. Podemos considerar entonces al grupo de transformaciones $\mathbf{O}(2)$ como un grupo de matrices con la operación de multiplicación.

Nótese que en las matrices de rotaciones y reflexiones, la matriz *inversa*, esto es, la matriz asociada a la transformación inversa, resultó ser muy parecida a la original; en las reflexiones es la misma y en las rotaciones se intercambian los elementos fuera de la diagonal (que difieren sólo en el signo). Esto no es coincidencia sino parte de algo mucho más general, que haremos explícito en esta sección y que será importante en capítulos posteriores.

Llamemos a una matriz *ortogonal* si es cuadrada y su función lineal asociada es una transformación ortogonal (que preserva producto interior). Entonces, si A es ortogonal, todas sus columnas son vectores unitarios (pues son la imagen de vectores unitarios, los canónicos) y además por parejas son ortogonales (pues así es la base canónica y esto se detecta por el producto interior). Dicho más explícitamente, si $A = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ entonces $\mathbf{u}_i = f(\mathbf{e}_i)$ donde f es su función asociada ($f(\mathbf{x}) = A\mathbf{x}$), y si $f \in \mathbf{O}(2)$ se cumple que

$$\mathbf{u}_i \cdot \mathbf{u}_j = \mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij} := \begin{cases} 1 & \text{si} \quad i = j \\ 0 & \text{si} \quad i \neq j, \end{cases}$$

donde δ_{ij} es conocida como "la delta de Kroenecker". En este caso, al conjunto ordenado de vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ se le llama *base ortonormal* (pues son vectores unitarios o normalizados y mutuamente ortogonales). Pero nótese que estos n^2 pro-

ductos interiores $(\mathbf{u}_i \cdot \mathbf{u}_j)$ son las entradas de la matriz

$$\begin{pmatrix} \mathbf{u}_1^\top \\ \mathbf{u}_2^\top \\ \vdots \\ \mathbf{u}_n^\top \end{pmatrix} (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = \begin{pmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \mathbf{u}_1 \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_1 \cdot \mathbf{u}_n \\ \mathbf{u}_2 \cdot \mathbf{u}_1 & \mathbf{u}_2 \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_2 \cdot \mathbf{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_n \cdot \mathbf{u}_1 & \mathbf{u}_n \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_n \cdot \mathbf{u}_n \end{pmatrix}$$

y que el que sean iguales a la delta de Kroenecker quiere decir que ésta es la matriz identidad. Así que si llamamos *transpuesta* de una matriz a la que se obtiene cambiando columnas por renglones, hemos demostrado que la "inversa" de una matriz ortogonal es su transpuesta.

En general, llamemos la *transpuesta* de una matriz $A = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ a la matriz

$$A^\top = \begin{pmatrix} \mathbf{v}_1^\top \\ \mathbf{v}_2^\top \\ \vdots \\ \mathbf{v}_n^\top \end{pmatrix}$$

que se obtiene al reflejar las entradas en la diagonal; si A es $m \times n$ entonces $A^\top$ es $n \times m$ y se generaliza nuestra noción previa de transpuestos de vectores. Hemos demostrado la mitad del siguiente teorema. Su otra mitad, para el caso que más nos interesa, $n = 2$, es consecuencia inmediata de la Proposición 3.12. Sin embargo damos una nueva demostración, muy general, para ejemplificar el poder de la notación matricial.

Teorema 3.17 *Una matriz A de $n \times n$ es ortogonal (la matriz de una transformación ortogonal) si y sólo si*

$$A^\top A = I.$$

Demostración. Hemos demostrado que si la función $f(\mathbf{x}) = A\mathbf{x}$ preserva producto interior entonces su matriz cumple que $A^\top A = I$. Supongamos ahora que $A^\top A = I$ y demostremos que su transformación asociada preserva producto interior. Para esto, expresaremos al producto interior como un caso más del producto de matrices. Para cualquier $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, se tiene

$$f(\mathbf{x}) \cdot f(\mathbf{y}) = A\mathbf{x} \cdot A\mathbf{y} = (A\mathbf{x})^\top A\mathbf{y} = \left(\mathbf{x}^\top A^\top\right) A\mathbf{y} = \mathbf{x}^\top \left(A^\top A\right) \mathbf{y} = \mathbf{x}^\top (I)\, \mathbf{y} = \mathbf{x} \cdot \mathbf{y},$$

donde hemos usado que $(A\mathbf{x})^\top = \mathbf{x}^\top A^\top$ (lo cual es un caso particular de la fórmula $(AB)^\top = B^\top A^\top$ que se deja como ejercicio). Por lo tanto, f es ortogonal (y también A lo es). □

Así que aunque las entradas de las matrices ortogonales sean complicadas, encontrar sus inversas es muy fácil. Y se explica por qué las inversas de las ortogonales 2×2 fueron tan parecidas a sí mismas.

EJERCICIO 3.85 Sea $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ una base ortonormal de $\mathbb{R}^3$, y sea $f : \mathbb{R}^3 \to \mathbb{R}^3$ definida por

$$f(x, y, z) = x\,\boldsymbol{u} + y\,\boldsymbol{v} + z\,\boldsymbol{w}.$$

Demuestra directamente que f es ortogonal (preserva producto interior). (Compara con la Proposición 3.12.)

EJERCICIO 3.86 Demuestra que $(A\boldsymbol{x})^\top = \boldsymbol{x}^\top A^\top$ cuando A es una matriz $m \times n$ y $\boldsymbol{x}$ un vector $(n \times 1)$.

EJERCICIO 3.87 Demuestra que $(AB)^\top = B^\top A^\top$ cuando A es una matriz $m \times n$ y B una matriz $(n \times k)$.

EJERCICIO 3.88 Demuestra que la "matriz de Pitágoras"

$$\frac{1}{5}\begin{pmatrix} 4 & 3 \\ -3 & 4 \end{pmatrix}$$

es ortogonal.

3.6 El Grupo General Lineal (GL(2))

En esta sección estudiamos las matrices cuyas funciones lineales asociadas son transformaciones, es decir, biyectivas. La definición abstracta será entonces muy fácil, aunque después habrá que hacerla más concreta.

Definición 3.6.1 Una matriz A de $n \times n$ es *invertible* si existe otra matriz de $n \times n$, llamada su *inversa* y denotada A^{-1} tal que $AA^{-1} = A^{-1}A = I$. Al conjunto de todas las matrices invertibles de $n \times n$ se le llama el *grupo general lineal de orden* n y se le denota $\mathbf{GL}(n)$. Así que $A \in \mathbf{GL}(n)$ quiere decir que A es invertible y de $n \times n$.

Para demostrar que las transformaciones lineales $f : \mathbb{R}^n \to \mathbb{R}^n$ tienen matrices asociadas que son invertibles sólo habría que ver que su función inversa también es lineal (ver el ejercicio siguiente), y entonces la matriz asociada a ésta será la inversa. Y al revés, si una matriz tiene inversa entonces las funciones lineales asociadas también son inversas. Así que a $\mathbf{GL}(2)$ se le puede pensar como grupo de transformaciones, tal y como lo hicimos en las primeras secciones de este capítulo, o bien como grupo de matrices (donde, hay que enfatizar, la operación del grupo es la multiplicación de matrices).

Ejemplos de matrices invertibles son las ortogonales que ya hemos visto. Así, tenemos que $\mathbf{O}(n)$ es un subgrupo de $\mathbf{GL}(n)$.

EJERCICIO 3.89 Sea $f : \mathbb{R}^n \to \mathbb{R}^n$ una transformación lineal; es decir, una función lineal con inversa f^{-1}. Demuestra que su inversa también es lineal.

3.6.1 El determinante

Veamos ahora el problema de detectar de manera sencilla las matrices invertibles de 2×2. Sea

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

una matriz cualquiera de 2×2. Consideremos el problema general de encontrar su inversa, y de paso ver cúando ésta existe. Tomemos entonces una matriz X cuyas entradas son incógnitas, digamos x, y, z, w, que cumplan

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x & z \\ y & w \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Esto nos da un sistema de cuatro ecuaciones con cuatro incógnitas, que en realidad se parte en los dos sistemas de dos ecuaciones con dos incógnitas siguientes

$$\begin{array}{c} a\,x + b\,y = 1 \\ c\,x + d\,y = 0 \end{array} \quad \text{y} \quad \begin{array}{c} a\,z + b\,w = 0 \\ c\,z + d\,w = 1, \end{array}$$

que provienen de cada una de las columnas de la identidad, o bien de las columnas de incógnitas. Recordando la Sección 1.6, o bien volviéndola a resolver, se obtiene que si $ad - bc \neq 0$, entonces los dos sistemas tienen soluciónes únicas que son precisamente

$$\begin{array}{rclcl} x & = & \dfrac{d}{ad - bc} & & z = \dfrac{-b}{ad - bc} \\ y & = & \dfrac{-c}{ad - bc} & & w = \dfrac{a}{ad - bc}. \end{array} \tag{3.13}$$

Para enunciar esto de manera elegante conviene definir el *determinante* de la matriz A como el número

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} := ad - bc.$$

Nótese que si $A = (\mathbf{u}, \mathbf{v})$, entonces $\det(A) = \det(\mathbf{u}, \mathbf{v})$ según lo definimos en la Sección 1.7, de tal manera que el determinante de una matriz es el determinante de sus columnas o el determinante de un sistema de ecuaciones asociado y sólo estamos ampliando naturalmente el espectro del término.

Teorema 3.18 *Sea* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$*, entonces* A *es invertible si y sólo si* $\det(A) = ad - bc \neq 0$*; y en este caso*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Hemos demostrado que si $\det(A) \neq 0$ entonces A es invertible y su inversa tiene la expresión indicada por (3.13). Nos falta ver que si A tiene inversa entonces su determinante no es cero. Pero esto será consecuencia inmediata del siguiente lema que nos da el determinante de un producto.

Lema 3.19 *Sean* A *y* B *matrices* 2×2*, entonces* $\det(AB) = \det(A)\det(B)$*.*

Demostración. Sea A como arriba y consideremos B con letras griegas. Se tiene entonces que

$$\begin{aligned}
\det(AB) &= \det\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\right) \\
&= \det\begin{pmatrix} a\alpha + b\gamma & a\beta + b\delta \\ c\alpha + d\gamma & c\beta + d\delta \end{pmatrix} \\
&= (a\alpha + b\gamma)(c\beta + d\delta) - (a\beta + b\delta)(c\alpha + d\gamma) \\
&= a\alpha c\beta + a\alpha d\delta + b\gamma c\beta + b\gamma d\delta \\
&\quad - a\beta c\alpha - a\beta d\gamma - b\delta c\alpha - b\delta d\gamma \\
&= ad(\alpha\delta - \beta\gamma) + bc(\beta\gamma - \alpha\delta) \\
&= (ad - bc)(\alpha\delta - \beta\gamma) = \det(A)\det(B).
\end{aligned}$$

□

Demostración. (del Teorema 3.18). Si A es una matriz invertible entonces existe A^{-1} tal que $AA^{-1} = I$. Por el lema anterior se tiene entonces que $\det(A)\det(A^{-1}) = \det(I) = 1$, y por lo tanto $\det(A)$ no puede ser 0. □

Veamos por último que el determinante de una matriz tiene un significado geométrico. Por el Teorema 1.30, corresponde al área dirigida del paralelogramo generado por sus vectores columna.

Teorema 3.20 *Sea* A *una matriz* 2×2*, entonces* $\det(A)$ *es el área del paralelogramo definido por sus vectores columna.* □

Se dice que la matriz A, o que la función $f(\mathbf{x}) = A\mathbf{x}$, *invierte orientación* si $\det(A) < 0$, o bien, que la *preserva* si $\det(A) > 0$. Esto corresponde a que cualquier figura que no tenga simetrías de espejo (por ejemplo, un "manquito" de la mano izquierda) se puede orientar, y bajo transformaciones lineales que preservan orientación, la figura puede deformarse pero conserva su orientación (el manquito seguirá siendo manquito de la mano izquierda); pero bajo transformaciones que invierten orientación se verá como manquito de la otra mano.



Pero además, el determinante de A representa lo que la función f "distorsiona" las áreas. Es decir, si una figura cualquiera F tiene área a, entonces $f(F)$ tendrá área $a\,|\det(A)|$. Sin pretender dar una demostración formal de esto, en general se puede argumentar como sigue. Puesto que la imagen del cuadrado unitario, C, bajo la función f asociada a la matriz $A = (\mathbf{u}, \mathbf{v})$, es el paralelogramo generado por $\mathbf{u}$ y $\mathbf{v}$, primero hay que ver que cuadraditos con lados paralelos a los ejes y de lado ε (y área ε^2) van a dar a paralelogramos de área $\varepsilon^2 \det(A)$; y luego observar que las áreas se pueden aproximar por cuadraditos disjuntos de este tipo. Con esta interpretación geométrica del determinante es natural que el determinante de un producto sea el producto de los determinantes, pues el factor de distorsión de áreas de una composición debía de ser el producto de factores de distorsión.

Observemos por último que el determinante distingue nuestras dos clases de matrices ortogonales (en **O** (2)); es 1 para las rotaciones, y -1 para las reflexiones. Es decir, las rotaciones preservan orientación y las reflexiones la invierten. Además su multiplicación corresponde a cómo se multiplican 1 y -1.

EJERCICIO 3.90 Demuestra que el determinante de las rotaciones es 1 y el de las reflexiones -1.

EJERCICIO 3.91 Sea $f : \mathbb{R}^2 \to \mathbb{R}^2$ la transformación lineal asociada a la matriz $A \in$ **GL**(2), y sea T cualquier triángulo en el plano. Demuestra que el área de $f(T)$ es $|\det(A)|$ por el área de T.

EJERCICIO 3.92 Sea $\mathbb{C} = \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \mid a, b \in \mathbb{R} \right\}$; y sea $\mathbb{C}^* = \mathbb{C} - \{0\}$, es decir, todo $\mathbb{C}$ menos la matriz 0. Describe las funciones lineales asociadas. Demuestra que $\mathbb{C}^*$ es un subgrupo de **GL**(2). Compara estas matrices con los números complejos.

EJERCICIO 3.93 Encuentra (en cada caso) la matriz de la transformación lineal $f : \mathbb{R}^2 \to \mathbb{R}^2$ que cumple:

a) $f(2,1) = (1,0)$ y $f(3,2) = (0,1)$
b) $f(2,5) = (1,0)$ y $f(1,2) = (0,1)$
c) $f(2,1) = (2,5)$ y $f(3,2) = (1,2)$,

donde hemos usado vectores renglón en vez de vectores columna por simplicidad.

EJERCICIO 3.94 Demuestra que una transformación lineal de $\mathbb{R}^2$ manda rectas en rectas.

EJERCICIO 3.95 Considera las rectas $(\ell_1 : x+y = 2)$; $(\ell_2 : 2x-y = 1)$; $(\ell_3 : 2x+y = 6)$ y $(\ell_4 : x - 4y = 3)$. Encuentra la transformación lineal $f : \mathbb{R}^2 \to \mathbb{R}^2$ que cumple $f(\ell_1) = \ell_3$ y $f(\ell_2) = \ell_4$.

3.6.2 Sistemas de ecuaciones lineales II

Vale la pena resumir y reinterpretar explícitamente lo que hemos hecho sobre sistemas de dos ecuaciones lineales en dos incógnitas con nuestro nuevo lenguaje de matrices y funciones. Un sistema tal es una ecuación

$$A\mathbf{x} = \mathbf{b},$$

donde A es una matriz 2×2 dada y $\mathbf{b}$ es un vector constante en $\mathbb{R}^2$. Si f es la función líneal asociada a la matriz, resolver el sistema es encontrar un vector $\mathbf{x}$ que bajo f vaya a $\mathbf{b}$. Si A tiene inversa, resolver el sistema es fácil pues se multiplican ambos lados por A^{-1} para obtener

$$\mathbf{x} = A^{-1}\mathbf{b},$$

que es justo la primera fórmula que obtuvimos en la Sección 1.6.1. Y podemos resumir lo que hemos hecho en diferentes lados con el siguiente teorema que esencialmente ya hemos demostrado.

Teorema 3.21 *Sea A una matriz de 2×2 y $f : \mathbb{R}^2 \to \mathbb{R}^2$ su función lineal asociada. Entonces son equivalentes:*
i) $\det(A) \neq 0$.
ii) La matriz A tiene inversa.
iii) El sistema $A\mathbf{x} = \mathbf{b}$ tiene solución para toda $\mathbf{b} \in \mathbb{R}^2$.
iv) El sistema $A\mathbf{x} = \mathbf{b}$ tiene solución única para alguna $\mathbf{b} \in \mathbb{R}^2$.
v) La función f es inyectiva.
vi) La función f es sobre.
vii) La función f es biyectiva. □

Además sabemos que si $\det(A) = 0$ entonces la imagen de f es una recta por el origen (o el origen cuando A es la matriz $\mathbf{0}$) y entonces para los puntos $\mathbf{b}$ fuera de esa recta, el sistema $A\mathbf{x} = \mathbf{b}$ no tiene solución, mientras que para los puntos en ella hay una recta de soluciones; además estas rectas de soluciones son paralelas.

3.7 Transformaciones afines

El papel que tuvieron las transformaciones ortogonales respecto de las isometrías lo tendrán ahora las transformaciones lineales respecto de las que llamaremos afines. Es decir, si componemos una transformación lineal con una traslación no trivial, ya no

obtenemos una lineal (pues el origen se mueve) sino algo que llamaremos transformación afín, generalizando para $n > 1$ a las transformaciones afines de $\mathbb{R}$, que ya estudiamos.

Definición 3.7.1 Una función $f : \mathbb{R}^n \to \mathbb{R}^n$ es una *transformación afín* si se escribe Dibujo

$$f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$$

para toda $\mathbf{x} \in \mathbb{R}^n$, donde $A \in \mathbf{GL}(n)$ y $\mathbf{b}$ es un vector fijo en $\mathbb{R}^n$. Dicho de otra manera, f es una *transformación afín* si existe una trasformación lineal $f_o : \mathbb{R}^n \to \mathbb{R}^n$ ($f_o(\mathbf{x}) = A\mathbf{x}$) y una translación $\tau_{\mathbf{b}}$ tales que

$$f = \tau_{\mathbf{b}} \circ f_o.$$

El conjunto de todas las transformaciones afines de $\mathbb{R}^n$ forman un grupo de transformaciones, llamado el *grupo afín de* $\mathbb{R}^n$ y denotado con $\mathbf{Af}(n)$.

Lo primero que hay que observar es que una transformación afín es efectivamente una transformación; y esto es claro pues es la composición de dos funciones biyectivas. Para ver que $\mathbf{Af}(n)$ es efectivamente un grupo, hay que ver que es cerrado bajo inversas y composición, pues la identidad (y de hecho cualquier transformación lineal) está en $\mathbf{Af}(n)$. Si $f \in \mathbf{Af}(n)$ (dada por la fórmula de la definición) para obtener la fórmula explícita de su inversa, despejaremos $\mathbf{x}$ en la ecuación

$$\mathbf{y} = A\mathbf{x} + \mathbf{b},$$

donde $\mathbf{y}$ desempeña el papel de $f(\mathbf{x})$. Pasando $\mathbf{b}$ del otro lado y luego multiplicando por A^{-1} (que existe pues $A \in \mathbf{GL}(n)$) se obtiene

$$\mathbf{x} = (A^{-1})\mathbf{y} - (A^{-1})\mathbf{b}.$$

Sea entonces $g : \mathbb{R}^n \to \mathbb{R}^n$ definida por la ecuación $g(\mathbf{x}) = (A^{-1})\mathbf{x} - (A^{-1})\mathbf{b}$ (nótese que $g \in \mathbf{Af}(n)$ por definición) y se tiene que

$$\begin{aligned} (g \circ f)(\mathbf{x}) &= g(A\mathbf{x} + \mathbf{b}) \\ &= (A^{-1})(A\mathbf{x} + \mathbf{b}) - (A^{-1})\mathbf{b} \\ &= \left(A^{-1}A\right)\mathbf{x} + A^{-1}\mathbf{b} - A^{-1}\mathbf{b} = \mathbf{x}, \end{aligned}$$

y análogamente se tiene que $(f \circ g)(\mathbf{x}) = \mathbf{x}$, así que g es la inversa de f.

Si tomamos ahora $g : \mathbb{R}^n \to \mathbb{R}^n$ como cualquier otra transformación afín, entonces se escribe $g(\mathbf{x}) = B\mathbf{x} + \mathbf{c}$ para algunos $B \in \mathbf{GL}(n)$ y $\mathbf{c} \in \mathbb{R}^n$ de tal forma que $(g \circ f)$ también es una transformación afín pues se escribe

$$(g \circ f)(\mathbf{x}) = B(A\mathbf{x} + \mathbf{b}) + \mathbf{c} = (BA)\mathbf{x} + (B\mathbf{b} + \mathbf{c}).$$

Por lo tanto, $\mathbf{Af}(n)$ sí es un grupo de transformaciones. Dibujo

Las transformaciones afines preservan rectas, es decir, mandan rectas en rectas. Pues si una recta ℓ está definida como

$$\ell = \{\mathbf{p} + t\mathbf{v} \mid t \in \mathbb{R}\}$$

y $f \in \mathbf{Af}(n)$, es como arriba, entonces

$$\begin{aligned} f(\mathbf{p} + t\mathbf{v}) &= A(\mathbf{p} + t\mathbf{v}) + \mathbf{b} \\ &= (A\mathbf{p} + \mathbf{b}) + t(A\mathbf{v}) \\ &= f(\mathbf{p}) + t(A\mathbf{v}), \end{aligned}$$

de tal manera que

$$f(\ell) \subseteq \{f(\mathbf{p}) + t(A\mathbf{v}) \mid t \in \mathbb{R}\}.$$

Y este último conjunto es la recta que pasa por $f(\mathbf{p})$ con dirección $A\mathbf{v}$ (nótese que $A\mathbf{v} \neq \mathbf{0}$, pues $\mathbf{v} \neq \mathbf{0}$ y A es invertible); así que más que una contención, la última es justo una igualdad (ambos conjuntos están parametrizados por $\mathbb{R}$).

Resulta que las transformaciones afines son precisamente las que preservan rectas. Pero la demostración de que si una transformación preserva rectas es afín es complicada y requiere argumentos de continuidad que no vienen al caso en este momento. Así que mejor nos enfocamos en otra cosa que sí preservan.

3.7.1 Combinaciones afines (el Teorema de 3 en 3)

Dibujo

Otra manera de ver que las transformaciones afines preservan rectas es usando coordenadas baricéntricas. Sea ℓ la recta que pasa por los puntos $\mathbf{p}$ y $\mathbf{q}$. Entonces ℓ se puede escribir como

$$\ell = \{s\mathbf{p} + t\mathbf{q} \mid s, t \in \mathbb{R}, \quad s + t = 1\}.$$

Veamos qué hace la transformación afín f al punto $s\mathbf{p} + t\mathbf{q}$ cuando $s + t = 1$:

$$\begin{aligned} f(s\mathbf{p} + t\mathbf{q}) &= A(s\mathbf{p} + t\mathbf{q}) + \mathbf{b} \\ &= s(A\mathbf{p}) + t(A\mathbf{q}) + (s + t)\,\mathbf{b} \\ &= s(A\mathbf{p} + \mathbf{b}) + t(A\mathbf{q} + \mathbf{b}) \\ &= s\,f(\mathbf{p}) + t\,f(\mathbf{q}), \end{aligned}$$

donde hemos usado que $\mathbf{b} = s\mathbf{b} + t\mathbf{b}$ pues $s + t = 1$. Así que la recta por $\mathbf{p}$ y $\mathbf{q}$ va, bajo f, a la recta por $f(\mathbf{p})$ y $f(\mathbf{q})$ y preserva las coordenadas baricéntricas. Y esto se puede generalizar.

Una combinación lineal $\sum_{i=0}^{k} \lambda_i \mathbf{v}_i$ se llama *combinación afín* si los coeficientes cumplen que $\sum_{i=0}^{k} \lambda_i = 1$. Le hemos llamado "combinación baricéntrica" cuando intervienen dos puntos distintos o tres no colineales, pero en el caso general no nos interesa qué propiedades tengan los puntos involucrados $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k$. Su importancia, entre otras cosas, es que las transformaciones afines las tratan como las lineales tratan a las combinaciones lineales:

Teorema 3.22 *Una transformación* $f : \mathbb{R}^n \to \mathbb{R}^n$ *es afín si y sólo si preserva combinaciones afines; es decir, si y sólo si*

$$f(\sum_{i=0}^{k} \lambda_i \mathbf{v}_i) = \sum_{i=0}^{k} \lambda_i f(\mathbf{v}_i) \quad \textit{cuando} \quad \sum_{i=0}^{k} \lambda_i = 1.$$

Demostración. Si $f : \mathbb{R}^n \to \mathbb{R}^n$ es afín, entonces se escribe $f(\mathbf{x}) = A\mathbf{x}+\mathbf{b}$ para alguna matriz A de $n \times n$ y un vector $\mathbf{b} \in \mathbb{R}^n$. Veamos que preserva combinaciones afines. Sean $\lambda_i \in \mathbb{R}$, $i = 0, 1, \dots, k$, tales que $\sum_{i=0}^{k} \lambda_i = 1$; y sean $\mathbf{v}_i \in \mathbb{R}^n$, $i = 0, 1, \dots, k$, vectores cualesquiera. Entonces

$$\begin{aligned} f(\sum_{i=0}^{k} \lambda_i \mathbf{v}_i) &= A(\sum_{i=0}^{k} \lambda_i \mathbf{v}_i) + \mathbf{b} = \sum_{i=0}^{k} \lambda_i (A\mathbf{v}_i) + \mathbf{b} \\ &= \sum_{i=0}^{k} \lambda_i (A\mathbf{v}_i) + \sum_{i=0}^{k} \lambda_i \mathbf{b} = \sum_{i=0}^{k} \lambda_i (A\mathbf{v}_i + \mathbf{b}) = \sum_{i=0}^{k} \lambda_i f(\mathbf{v}_i). \end{aligned}$$

Donde se usó la linearidad en la segunda igualdad, y luego fue crucial que $\sum_{i=0}^{k} \lambda_i = 1$ para poder "distribuir la $\mathbf{b}$" en la sumatoria.

Por el otro lado, si nos dan f que preserva combinaciones afines, debemos encontrar la matriz A (o su función lineal asociada) y el vector $\mathbf{b}$ que la definan. Encontrar el vector $\mathbf{b}$ es fácil pues es dónde cualquier transformación afín manda el origen; definamos entonces

$$\mathbf{b} := f(\mathbf{0}).$$

Y, como sucedió con las isometrías, debemos definir la función $g : \mathbb{R}^n \to \mathbb{R}^n$ con

$$g(\mathbf{x}) := f(\mathbf{x}) - \mathbf{b}$$

y demostrar que es lineal para concluir la demostración; pues entonces $f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{b}$. Veamos entonces que g preserva combinaciones lineales.

Sean $\mathbf{v}_1, \dots, \mathbf{v}_k$ vectores en $\mathbb{R}^n$ y $\lambda_1, \dots, \lambda_n$ coeficientes (en $\mathbb{R}$) arbitrarios. Debemos demostrar que $g\left(\sum_{i=1}^{k} \lambda_i \mathbf{v}_i\right) = \sum_{i=1}^{k} \lambda_i g(\mathbf{v}_i)$, pero nuestra hipótesis sólo nos dice cómo se comporta f respecto a combinaciones afines. Podemos convertir esta combinación lineal en una afín usando el vector $\mathbf{0}$ como comodín. Sean

$$\mathbf{v}_0 := \mathbf{0} \quad \text{y} \quad \lambda_0 := 1 - \sum_{i=1}^{k} \lambda_i$$

de tal manera que $\sum_{i=0}^{k} \lambda_i = 1$ (ojo, empezamos la sumatoria en 0) y $\sum_{i=0}^{k} \lambda_i \mathbf{v}_i =$

$\sum_{i=1}^{k}\lambda_i \mathbf{v}_i$; y ahora sí podemos usar nuestra hipótesis para obtener

$$\begin{aligned} g\left(\sum_{i=1}^{k}\lambda_i\mathbf{v}_i\right) &= f\left(\sum_{i=1}^{k}\lambda_i\mathbf{v}_i\right)-\mathbf{b} = f\left(\sum_{i=0}^{k}\lambda_i\mathbf{v}_i\right)-\left(\sum_{i=0}^{k}\lambda_i\right)\mathbf{b} \\ &= \sum_{i=0}^{k}\lambda_i f(\mathbf{v}_i)-\sum_{i=0}^{k}\lambda_i\mathbf{b} = \sum_{i=0}^{k}\lambda_i\left(f(\mathbf{v}_i)-\mathbf{b}\right) \\ &= \sum_{i=0}^{k}\lambda_i g(\mathbf{v}_i) = \lambda_0 g(\mathbf{v}_0)+\sum_{i=1}^{k}\lambda_i g(\mathbf{v}_i) = \sum_{i=1}^{k}\lambda_i g(\mathbf{v}_i), \end{aligned}$$

pues $g(\mathbf{v}_0) = g(\mathbf{0}) = \mathbf{0}$. Lo cual concluye la demostración del teorema. □

Vale la pena notar que el concepto de transformación afín se puede generalizar al de función afín cuando la matriz A no es necesariamente invertible; una *función afín* será entonces una función lineal seguida de una traslación; ni siquiera necesitamos pedir que el dominio y el codominio coincidan. Si se revisa con cuidado el teorema anterior, se notará que vale en esta generalidad. Pero no lo hicimos así pues nuestro interes principal está en las transformaciones.

Dibujo

En el caso que más nos interesa, $n = 2$, este teorema y su demostración tienen como corolario que una transformación afín de $\mathbb{R}^2$ queda determinada por sus valores en el *triángulo canónico* con vértices $\mathbf{e}_0 := \mathbf{0}, \mathbf{e}_1, \mathbf{e}_2$, y que estos pueden ir a cualquier otro triángulo. Explícitamente, dado un triángulo T con vértices $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2$, la transformación afín que manda al triángulo canónico en T es

$$f(\mathbf{x}) = (\mathbf{a}_1 - \mathbf{a}_0, \mathbf{a}_2 - \mathbf{a}_0)\mathbf{x} + \mathbf{a}_0,$$

que cumple claramente que $f(\mathbf{e}_i) = \mathbf{a}_i$. Hay que observar que esta función está bien definida independientemente de quiénes sean $\mathbf{a}_0, \mathbf{a}_1$ y $\mathbf{a}_2$, pues tanto la matriz 2×2 como el vector por el que se traslada lo están. Pero que es una transformación sólo cuando $\mathbf{a}_0, \mathbf{a}_1$ y $\mathbf{a}_2$ no son colineales (que es lo que entendemos como *triángulo*), pues entonces la matriz es invertible.

Y como corolario de este hecho obtenemos el siguiente teorema.

Dibujo

Teorema 3.23 (3 en 3) *Dados dos triángulos con vértices* $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2$ *y* $\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2$ *respectivamente, existe una única transformación afín* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *que cumple* $f(\mathbf{a}_i) = \mathbf{b}_i$ *para* $i = 0, 1, 2$.

Demostración. Basta componer la inversa de la transformación que manda al triángulo canónico en el $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2$ con la que lo manda al $\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2$. La unicidad se sigue de la unicidad de las dos transformaciones usadas. □

Un ejemplo

Para fijar ideas veamos un ejemplo numérico. Sean

$$\begin{aligned} \mathbf{a}_0 &= \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad \mathbf{a}_1 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}; \\ \mathbf{b}_0 &= \begin{pmatrix} -5 \\ 6 \end{pmatrix}, \quad \mathbf{b}_1 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} -4 \\ 1 \end{pmatrix}. \end{aligned}$$

Queremos encontrar la transformación afín $f : \mathbb{R}^2 \to \mathbb{R}^2$ que cumple $f(\mathbf{a}_i) = \mathbf{b}_i$ para $i = 0, 1, 2$. Dibujos

Un método infalible (si no cometemos un error en el camino) es seguir la demostración del último teorema (el 6.4). Es decir, la transformación que manda al canónico en el triángulo de las $\mathbf{a}$'s es

$$\begin{pmatrix} 3 & -1 \\ -3 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

luego habrá que encontrar su inversa y componerla con la que manda al canónico en el triángulo de las $\mathbf{b}$'s. Pero dejaremos este método para que el lector lo siga paso a paso y compruebe el resultado con una solución diferente que es la que aquí seguiremos.

Ataquemos el problema directamente. Estamos buscando una matriz

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \quad \text{y un vector} \quad \mathbf{b} = \begin{pmatrix} \lambda \\ \mu \end{pmatrix}$$

para definir la función $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ que cumpla $f(\mathbf{a}_i) = \mathbf{b}_i$ para $i = 0, 1, 2$. Como las $\mathbf{a}$'s y las $\mathbf{b}$'s tienen valores numéricos explícitos, estas condiciones nos darán ecuaciones explícitas con las seis incógnitas $\alpha, \beta, \gamma, \delta, \lambda, \mu$ que podemos resolver directamente. La condición $f(\mathbf{a}_0) = \mathbf{b}_0$ nos da

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} + \begin{pmatrix} \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} -5 \\ 6 \end{pmatrix},$$

que se traduce en una ecuación para cada coordenada, a saber:

$$\begin{aligned} 0\alpha + 2\beta + \lambda &= -5 \\ 0\gamma + 2\delta + \mu &= 6. \end{aligned}$$

Análogamente, las condiciones en las otras dos parejas de puntos nos dan las siguientes ecuaciones:

$$\begin{aligned} 3\alpha - \beta + \lambda &= 4 \\ 3\gamma - \delta + \mu &= 3 \\ -\alpha + \beta + \lambda &= -4 \\ -\gamma + \delta + \mu &= 1 \end{aligned}$$

Resulta que en vez de seis ecuaciones con seis incógnitas, tenemos dos sistemas de tres ecuaciones con tres incógnitas, y además los dos sistemas tienen los mismos coeficientes y sólo se diferencian por las constantes. De hecho los podemos reescribir como

$$A\begin{pmatrix}\alpha\\ \beta\\ \lambda\end{pmatrix}=\begin{pmatrix}-5\\ 4\\ -4\end{pmatrix}\quad \text{y}\quad A\begin{pmatrix}\gamma\\ \delta\\ \mu\end{pmatrix}=\begin{pmatrix}6\\ 3\\ 1\end{pmatrix},$$

$$\text{donde}\qquad A=\begin{pmatrix}0 & 2 & 1\\ 3 & -1 & 1\\ -1 & 1 & 1\end{pmatrix}.$$

Observemos que al resolver los dos sistemas de ecuaciones con el método de eliminar variables sumando múltiplos de una ecuación a otra, estaremos realizando las mismas operaciones, y éstas dependen nada más de los coeficientes, no del nombre de las incógnitas. Así que la manera de resolver los dos sistemas al mismo tiempo será añadiendo dos columnas a la matriz A que correspondan a las constantes de los dos sistemas, y luego sumando múltiplos de renglones a otros para llevar la parte 3×3 de la matriz a una de permutaciones que nos diga el valor explícito de todas las incógnitas. Hagámoslo. La matriz con que empezamos es

$$\left(\begin{array}{ccc:c:c}0 & 2 & 1 & -5 & 6\\ 3 & -1 & 1 & 4 & 3\\ -1 & 1 & 1 & -4 & 1\end{array}\right).$$

Sumándole 3 por el tercer renglón al segundo (para hacer 0 su primer entrada), operación que podemos denotar con $3\mathbf{r}_3\vec{+}\mathbf{r}_2$ y que equivale a sumar un múltiplo de una ecuación a otra, obtenemos

$$\xrightarrow{3\mathbf{r}_3\vec{+}\mathbf{r}_2}\begin{pmatrix}0 & 2 & 1 & -5 & 6\\ 0 & 2 & 4 & -8 & 6\\ -1 & 1 & 1 & -4 & 1\end{pmatrix}\xrightarrow{-\mathbf{r}_2\vec{+}\mathbf{r}_1}\begin{pmatrix}0 & 0 & -3 & 3 & 0\\ 0 & 2 & 4 & -8 & 6\\ 1 & -1 & -1 & 4 & -1\end{pmatrix}$$

$$\xrightarrow{(1/2)\mathbf{r}_2;-(1/3)\mathbf{r}_1}\begin{pmatrix}0 & 0 & 1 & -1 & 0\\ 0 & 1 & 2 & -4 & 3\\ 1 & -1 & -1 & 4 & -1\end{pmatrix}\xrightarrow{\mathbf{r}_2\vec{+}\mathbf{r}_3}\begin{pmatrix}0 & 0 & 1 & -1 & 0\\ 0 & 1 & 2 & -4 & 3\\ 1 & 0 & 1 & 0 & 2\end{pmatrix}$$

$$\xrightarrow{-2\mathbf{r}_1\vec{+}\mathbf{r}_2;-\mathbf{r}_1\vec{+}\mathbf{r}_3}\begin{pmatrix}0 & 0 & 1 & -1 & 0\\ 0 & 1 & 0 & -2 & 3\\ 1 & 0 & 0 & 1 & 2\end{pmatrix}.$$

Ahora, recordando nuestras incógnitas, el primer sistema nos dice que $\lambda=-1$, $\beta=-2$ y $\alpha=1$ y el segundo sistema que $\mu=0$, $\delta=3$ y $\gamma=2$, lo cual lo resuelve por completo. En resumen, la función que buscábamos es

$$f(\mathbf{x})=\begin{pmatrix}1 & -2\\ 2 & 3\end{pmatrix}\mathbf{x}+\begin{pmatrix}-1\\ 0\end{pmatrix}$$

y es fácil comprobarlo; ya sea revisando los valores que se pedían o con el otro método.

EJERCICIO 3.96 Para cada caso, encuentra la transformación afín que manda $\mathbf{a}_i$ en $\mathbf{b}_i$

a) $\mathbf{a}_0 = (1,2)^\top, \mathbf{a}_1 = (1,-1)^\top, \mathbf{a}_2 = (0,-1)^\top$ y $\mathbf{b}_0 = (1,4)^\top, \mathbf{b}_1 = (4,4)^\top, \mathbf{b}_2 = (3,3)^\top$
b) $\mathbf{a}_0 = (1,2)^\top, \mathbf{a}_1 = (0,1)^\top, \mathbf{a}_2 = (3,-1)^\top$ y $\mathbf{b}_0 = (5,3)^\top, \mathbf{b}_1 = (2,0)^\top, \mathbf{b}_2 = (6,-1)^\top$
c) $\mathbf{a}_0 = (1,2)^\top, \mathbf{a}_1 = (0,1)^\top, \mathbf{a}_2 = (3,-1)^\top$ y $\mathbf{b}_0 = (10,5)^\top, \mathbf{b}_1 = (3,3)^\top, \mathbf{b}_2 = (-4,4)^\top$

(¿Verdad que conviene relajar la notación sobre vectores?)

3.8 Isometrías II

Ya con la herramienta técnica que nos dan las matrices y en el contexto general que nos dan las transformaciones afines, regresamos brevemente a estudiar isometrías del plano y ver algunas cosas que se nos quedaron pendientes. Ahora sabemos que una isometría de $\mathbb{R}^2$ es una transformación que se escribe

$$f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}, \tag{3.14}$$

donde la matriz A es ortogonal ($A^\top = A^{-1}$) y es llamada la *parte lineal* de la isometría. Hay de dos tipos. Las que preservan orientación (con $\det(A) = 1$) y las que invierten orientación ($\det(A) = -1$). Para las primeras, la matriz A es una rotación en el origen y para las segundas, una reflexión en una recta por el origen. Puesto que al componer isometrías (o, en general, transformaciones afines) la parte lineal se obtiene componiendo las partes lineales, entonces las isometrías (o las afines) que preservan orientación forman un grupo, denotado con $\mathbf{Iso}^+(2)$ (respectivamente, $\mathbf{Af}^+(2)$). Ojo, las que invierten orientación **no** forman un grupo, ni siquiera contienen a la identidad y la composición de dos de ellas preserva orientación.

Lo primero que queremos ver es que las isometrías que preservan orientación son rotaciones en algún punto o traslaciones.

3.8.1 Rotaciones y traslaciones

Recordemos que definimos la rotación $\rho_{\theta,\mathbf{c}}$ de un ángulo θ con centro $\mathbf{c}$, conjugando con la traslación de $\mathbf{c}$ al origen, es decir, como $\tau_{\mathbf{c}} \circ \rho_\theta \circ \tau_{-\mathbf{c}}$, donde $\tau_{\mathbf{y}}$ es la traslación por $\mathbf{y}$ y ρ_θ es rotar un ángulo θ en el origen. Así que, usando matrices, la fórmula para $\rho_{\theta,\mathbf{c}}$ es

$$\begin{aligned} \rho_{\theta,\mathbf{c}}(\mathbf{x}) &= R_\theta(\mathbf{x} - \mathbf{c}) + \mathbf{c} \\ &= R_\theta\mathbf{x} + (\mathbf{c} - R_\theta\mathbf{c}). \end{aligned} \tag{3.15}$$

Está claro entonces que $\rho_{\theta,\mathbf{c}} \in \mathbf{Iso}^+(2)$, pues se escribe como en (3.14) con una $\mathbf{b}$ complicada, pero constante al fin. Y al revés, aunque cierto, no está tan claro. Para ver que $f \in \mathbf{Iso}^+(2)$ dada como en (3.14) es la rotación en algún centro, hay que encontrar un punto fijo, es decir, un punto $\mathbf{c}$ para el cual $f(\mathbf{c}) = \mathbf{c}$ (pues a su centro no lo mueve la rotación); que nos da la ecuación en $\mathbf{c}$

$$\begin{aligned} \mathbf{c} &= A\mathbf{c} + \mathbf{b} \\ \mathbf{c} - A\mathbf{c} &= \mathbf{b}. \end{aligned}$$

Esto coincide con la expresión (3.15), donde se usa el centro de rotación y nos da la traslación $\mathbf{b}$ en términos de él. Entonces, el problema de encontrar un punto fijo para la transformación (3.14) equivale a encontrar una solución a la ecuación

$$\mathbf{x} - A\mathbf{x} = \mathbf{b}$$

que en realidad es un sistema de ecuaciones. Y como tal, se puede reescribir

$$\begin{aligned} I\mathbf{x} - A\mathbf{x} &= \mathbf{b} \\ (I - A)\,\mathbf{x} &= \mathbf{b}, \end{aligned} \tag{3.16}$$

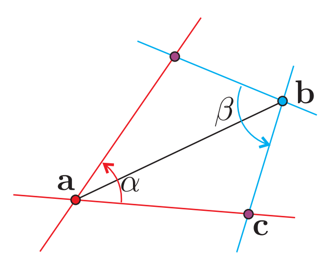
donde nos ha sido muy útil la suma de matrices. Sabemos que este sistema tiene solución única si y sólo si su determinante es distinto de cero. Para el caso que nos ocupa, cuando A es una matriz de rotación, por un ángulo θ digamos, se tiene

$$\begin{aligned} \det(I - R_\theta) &= \det \begin{pmatrix} 1 - \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & 1 - \cos\theta \end{pmatrix} \\ &= (1 - \cos\theta)^2 + \operatorname{sen}^2\theta \\ &= 1 - 2\cos\theta + \cos^2\theta + \operatorname{sen}^2\theta = 2(1 - \cos\theta). \end{aligned}$$

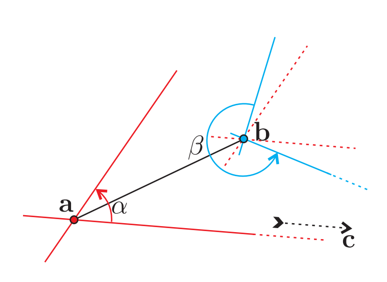
Podemos concluir que si $\theta \neq 0$ entonces $\det(I - R_\theta) \neq 0$ y tenemos una solución única del sistema (3.16) con $A = R_\theta$, que es el centro de rotación. Así que si $f \in \mathbf{Iso}^+(2)$, entonces es una rotación (alrededor de un punto) o bien una traslación (cuando el ángulo θ es cero en la discusión anterior). Además, hemos demostrado que una rotación (por un ángulo no cero, se sobreentiende) no tiene puntos fijos además de su centro.

Puesto que una rotación o una traslación se puede obtener *moviendo* continuamente (poco a poco) el plano, a veces se le llama a $\mathbf{Iso}^+(2)$ el grupo de *movimientos rígidos del plano.*

Como corolario se obtiene que la composición de dos rotaciones es una nueva rotación. El nuevo centro de rotación se puede obtener analíticamente resolviendo el correspondiente sistema de ecuaciones, pero lo veremos geométricamente.



Sean $\rho_{\alpha,\mathbf{a}}$ y $\rho_{\beta,\mathbf{b}}$ las rotaciones de ángulos α y β y centros $\mathbf{a}$ y $\mathbf{b}$ respectivamente. La composición de estas dos rotaciones será una rotación de un ángulo $\alpha+\beta$ (pues las partes lineales se componen) pero su centro depende del orden en que se componen. Para obtener el centro de la rotación $\rho_{\beta,\mathbf{b}}\circ\rho_{\alpha,\mathbf{a}}$, se traza el segmento de $\mathbf{a}$ a $\mathbf{b}$, la línea por $\mathbf{a}$ con ángulo $-\alpha/2$ respecto de este y la línea por $\mathbf{b}$ con ángulo $\beta/2$; la intersección $\mathbf{c}$ de estas dos líneas es el centro de rotación. Para verlo, persígase este punto según las rotaciones. Nótese que su reflejado respecto de la línea por $\mathbf{a}$ y $\mathbf{b}$ es el centro de rotación de la composición en el otro orden, $\rho_{\alpha,\mathbf{a}}\circ\rho_{\beta,\mathbf{b}}$.



Para que nuestra construcción de $\mathbf{c}$ tuviera sentido se necesita que las dos rectas cuya intersección lo definen sean no paralelas y éste es justo el caso cuando $\alpha+\beta\neq 0$. Pero nos podemos aproximar al valor $\beta=-\alpha$ poco a poco, y en este proceso el punto de intersección $\mathbf{c}$ se "va al infinito", así que las traslaciones son, en cierta manera, "rotaciones con centro al infinito".

EJERCICIO 3.97 Las *homotecias* son transformaciones de la forma $f(\mathbf{x})=k\mathbf{x}+\mathbf{b}$, donde a $k\neq 0$ se le llama el *factor de expansión* . Demuestra que las homotecias con factor de expansión no trivial (distinto de 1) tienen un *centro de expansión,* es decir, un punto fijo .

3.8.2 Reflexiones y "pasos"

Consideremos ahora las isometrías que invierten orientación. Se escriben como

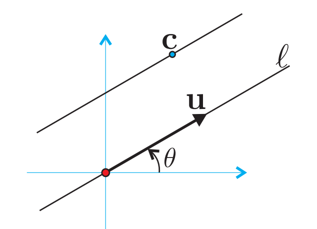
$$f(\mathbf{x})=E_\theta\mathbf{x}+\mathbf{b},$$

donde E_θ es una matriz de reflexión. Veamos primero si tiene puntos fijos, para lo cual hay que resolver el correspondiente sistema (3.16). En este caso el determinante es

$$\begin{aligned}\det(I-E_\theta) &= \det\begin{pmatrix}1-\cos 2\theta & -\operatorname{sen}2\theta\\ -\operatorname{sen}2\theta & 1+\cos 2\theta\end{pmatrix}\\ &= (1-\cos 2)(1+\cos 2\theta)-\operatorname{sen}^2 2\theta\\ &= 1-\cos^2 2\theta-\operatorname{sen}^2 2\theta = 1-1=0.\end{aligned}$$

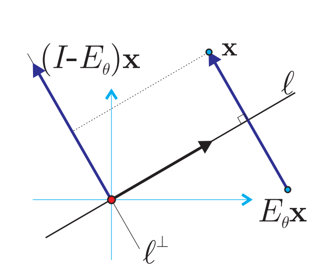
Y por lo tanto no tiene solución única: o no tiene solución o bien tiene muchas; y veremos que ambos casos son posibles.

Para $\mathbf{b} = 0$ tenemos que f es una reflexión y entonces las soluciones debieran ser los puntos de la recta espejo (a un ángulo θ con el eje x positivo), que llamaremos ℓ. Veamos: sea $\mathbf{u}$ el vector unitario que genera a ℓ, es decir, $\mathbf{u} = (\cos\theta, \operatorname{sen}\theta)$. Entonces

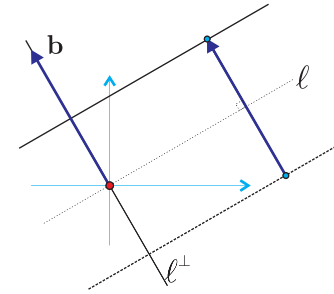


$$\begin{aligned} E_\theta \mathbf{u} &= \begin{pmatrix} \cos 2\theta & \operatorname{sen} 2\theta \\ \operatorname{sen} 2\theta & -\cos 2\theta \end{pmatrix} \begin{pmatrix} \cos\theta \\ \operatorname{sen}\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos 2\theta \cos\theta + \operatorname{sen} 2\theta \operatorname{sen}\theta \\ \operatorname{sen} 2\theta \cos\theta - \cos 2\theta \operatorname{sen}\theta \end{pmatrix} = \begin{pmatrix} \cos(2\theta - \theta) \\ \operatorname{sen}(2\theta - \theta) \end{pmatrix} = \mathbf{u}, \end{aligned}$$

donde hemos usado las fórmulas trigonométricas para la diferencia de ángulos (3.12). Y por lo tanto $E_\theta(t\mathbf{u}) = t(E_\theta \mathbf{u}) = t\mathbf{u}$; que nos dice que todos los puntos de la recta ℓ satisfacen la ecuación homogénea $(I - E_\theta)\mathbf{x} = 0$. Ahora bien, si para cierta $\mathbf{b}$, el sistema $(I - E_\theta)\mathbf{x} = \mathbf{b}$ tiene una solución particular, $\mathbf{c}$ digamos, entonces toda la recta $\ell + \mathbf{c}$, paralela a ℓ que pasa por $\mathbf{c}$, consta de soluciones para el sistema; es una reflexión con espejo $\ell + \mathbf{c}$. Pero nos falta determinar para cuáles $\mathbf{b}$ hay solución.

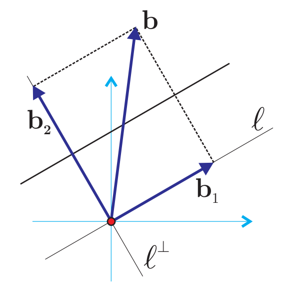


Para esto, conviene pensar geométricamente. Para cualquier $\mathbf{x} \in \mathbb{R}^2$, $(I - E_\theta)\mathbf{x} = \mathbf{x} - E_\theta\mathbf{x}$ es el vector que va de $E_\theta\mathbf{x}$ a $\mathbf{x}$, y como E_θ es una reflexión debe ser perpendicular al espejo; es más, debe ser el doble del vector que va de ℓ a $\mathbf{x}$ y es perpendicular a ℓ (recuérdese que así definimos reflexiones por primera vez). De tal manera que $(I - E_\theta)$ es, como función, la proyección ortogonal a $\ell^\perp : \mathbf{u} \cdot \mathbf{x} = 0$ seguida de la homotecia "multiplicar por 2". La imagen de la función $(I - E_\theta)$ es entonces $\ell^\perp$, y por lo tanto, sólo cuando $\mathbf{b} \in \ell^\perp$ la isometría $f(\mathbf{x}) = E_\theta\mathbf{x} + \mathbf{b}$ tiene puntos fijos. Más concretamente, si $\mathbf{u} \cdot \mathbf{b} = 0$, entonces reflejar en ℓ y luego trasladar por $\mathbf{b}$, regresa de la recta $\ell + (1/2)\mathbf{b}$ a su lugar (y punto a punto).



En el caso general, podemos expresar a cualquier $\mathbf{b} \in \mathbb{R}^2$ como una suma de sus *componentes* respecto a la base ortonormal $\mathbf{u}, \mathbf{u}^\perp$, es decir como

$$\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2,$$

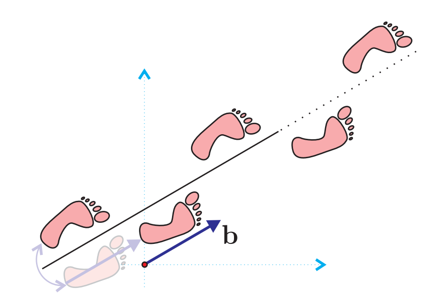


con $\mathbf{b}_1 \in \ell$ y $\mathbf{b}_2 \in \ell^\perp$ (de hecho, por el teorema de bases ortonormales: $\mathbf{b}_1 := (\mathbf{b} \cdot \mathbf{u})\,\mathbf{u}$ y $\mathbf{b}_2 := \left(\mathbf{b} \cdot \mathbf{u}^\perp\right)\mathbf{u}^\perp$). Y entonces la transformación $f(\mathbf{x}) = E_\theta\mathbf{x} + \mathbf{b}$ se puede escribir como

$$f(\mathbf{x}) = (E_\theta\mathbf{x} + \mathbf{b}_2) + \mathbf{b}_1,$$

que es una reflexión en la recta $\ell + (1/2)\,\mathbf{b}_2$ (por el caso anterior) y que en la fórmula es $(E_\theta \mathbf{x} + \mathbf{b}_2)$, seguida de una traslación $(\ldots + \mathbf{b}_1)$ en la dirección del espejo. Nótese que si $\mathbf{b}_1 \neq \mathbf{0}$, entonces f no tiene puntos fijos, los puntos de un lado del espejo van al otro lado y los que están en el espejo se trasladan en él, aunque el espejo $\ell + (1/2)\,\mathbf{b}_2$ como línea sí se queda en su lugar (todos sus puntos se mueven dentro de ella) y éste es entonces el caso en que nuestro sistema de ecuaciones que detecta puntos fijos no tiene solución.

A una isometría tal, una reflexión seguida de una traslación no trivial $(\mathbf{b} \neq \mathbf{0})$ en la dirección de la línea espejo, la llamaremos un *paso* pues genera la simetría que tiene un caminito de huellas (infinito) y corresponde al acto de dar un paso: cambiar de pie (reflejar) y adelantarlo un poco (trasladar). Si la aplicamos dos veces, nos da una traslación (por $2\mathbf{b}$), y si la seguimos aplicando (a ella y a su inversa) nos da un subgrupo de isometrías que son las simetrías de un camino infinito. Pero en fin, hemos demostrado:



Teorema 3.24 *Una isometría que invierte orientación es un paso o una reflexión (paso con traslación trivial).* □

EJERCICIO 3.98 Con la notación de los párrafos anteriores, demuestra con coordenadas que $(I - E_\theta)\,\mathbf{u}^\perp = 2\mathbf{u}^\perp$. (Tienes que usar las identidades trigonométricas para el doble del ángulo.)

EJERCICIO 3.99 Demuestra que si f es una isometría que invierte orientación, entonces $f^2 = f \circ f$ es una traslación.

EJERCICIO 3.100 Sea $f(\mathbf{x}) = E_\theta \mathbf{x} + \mathbf{b}$. Encuentra la expresión para f^{-1} y arguméntala geométricamente.

EJERCICIO 3.101 Demuestra analíticamente que la composición de dos reflexiones en lineas paralelas es una traslación en la dirección ortogonal a sus espejos.

3.8.3 Homotecias y semejanzas

El clásico *zoom* que aumenta (o disminuye) de tamaño una figura es lo que llamaremos una *homotecia.* Siempre tiene un centro de expansión, que cuando es el origen resulta ser una transformación lineal con matriz asociada

$$k\,I = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix},$$

donde $k > 0$; para $k > 1$ magnifica (es un *zoom* estrictamente hablando) y para $k < 1$ disminuye. Si la componemos con una traslación, por $\mathbf{b} \in \mathbb{R}^2$ digamos, se obtiene

una *homotecia* de *factor* k. Su centro de expansión $\mathbf{c}$ es el punto que se queda fijo y que se obtiene entonces resolviendo la ecuación

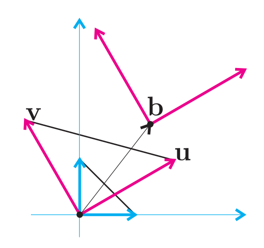
$$k\mathbf{x} + \mathbf{b} = \mathbf{x}.$$

Una *semejanza* es una transformación afín que preserva ángulos. Claramente las homotecias y las isometrías son semejanzas. Veremos que estas últimas se obtienen simplemente al dejar interactuar a las dos primeras.

Teorema 3.25 *Si* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *es una semejanza, entonces existen* $k > 0$, $A \in \boldsymbol{O}(2)$ *y* $\mathbf{b} \in \mathbb{R}^2$ *tales que*

$$f(\mathbf{x}) = k A\mathbf{x} + \mathbf{b}.$$

Demostración. Consideremos la transformación lineal $g(\mathbf{x}) = f(\mathbf{x}) - \mathbf{b}$, donde $\mathbf{b} := f(\mathbf{0})$, y notemos que preserva ángulos pues las traslaciones lo hacen. Sea $B = (\mathbf{u}, \mathbf{v})$ la matriz asociada a g. Como los vectores canónicos son ortogonales entonces $\mathbf{u}$ y $\mathbf{v}$ (sus imágenes bajo g) también lo son. Además tienen la misma norma pues si no fuera así, los ángulos chiquitos del triángulo $\mathbf{0}, \mathbf{u}, \mathbf{v}$ no medirían $\pi/4$ que son los del triángulo canónico. Sean $k = |\mathbf{u}| = |\mathbf{v}|$ y



$$A = \frac{1}{k}B.$$

Se tiene que $A \in \mathbf{O}(2)$, pues sus columnas son ortonormales, y además

$$f(\mathbf{x}) = g(\mathbf{x}) + \mathbf{b} = B\mathbf{x} + \mathbf{b} = kA\mathbf{x} + \mathbf{b}.$$

□

EJERCICIO 3.102 Encuentra la expresión de la homotecia de factor k y centro $\mathbf{c}$.

EJERCICIO 3.103 Demuestra que una transformación $f : \mathbb{R}^2 \to \mathbb{R}^2$ es una semejanza si y sólo si existe $k > 0$ tal que $d(f(\mathbf{x}), f(\mathbf{y})) = k\, d(\mathbf{x}, \mathbf{y})$ para todo $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$.

3.9 *Simetría plana

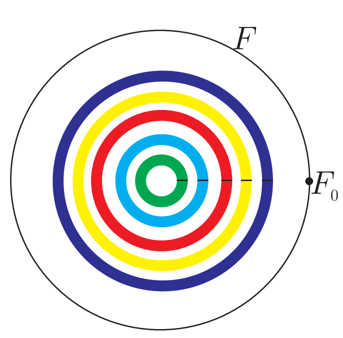
En la Sección 3.3.2 se definió el grupo de simetrías de una figura plana; en esta última sección estudiamos diferentes instancias de cuales pueden ser estos grupos y cómo son sus figuras asociadas.

3.9.1 El Teorema de Leonardo

Versión topológica

Por una *figura acotada* F entendemos un subconjunto del plano que a su vez está contenido en el interior de algún círculo; y por medio de una homotecia, podemos suponer que está dibujada en una hoja de papel. El problema general que nos guía es qué podemos decir sobre su *grupo de simetrías*

$$\mathbf{Sim}\,(\mathsf{F}) = \{\mathsf{f} \in \mathbf{Iso}\,(2) \mid \mathsf{f}\,(\mathsf{F}) = \mathsf{F}\}\,.$$



O bien, ¿qué grupos son de este tipo? Además de los grupos cíclicos y diédricos que vimos en la Sección 3.3.2 tenemos otros. Por ejemplo, si F es un círculo centrado en el origen, entonces $\mathbf{Sim}\,(\mathsf{F}) = \mathbf{O}\,(2)$. Nótese que también un conjunto de anillos concéntricos —un tiro al blanco— tiene este mismo grupo de simetrías. La figura F se genera por un pedacito F_0 –un simple punto para el círculo y para el tiro al blanco un intervalo por cada anillo– juntando todas las copias $\mathsf{f}(\mathsf{F}_0)$ al correr $\mathsf{f} \in \mathbf{O}\,(2)$; es decir, dejando que el grupo "barra" o "actúe" en una "figura fundamental" F_0. Y como en este caso el grupo es "continuo" (los ángulos de las rotaciones son un continuo) la figura que se genera también es "continua". Podemos pensar que la figura F se obtiene "poniendo tinta" (de diferentes colores para el tiro al blanco) en los puntos de F_0 y luego dejando que el grupo mueva a F_0 y vaya dejando huella. Todas las figuras que tienen simetría se deben obtener así, una porción de ella y el grupo la definen: "simetría, se dice coloquialmente, es una relación armoniosa entre las partes y el todo".

Otro ejemplo es tomar el grupo generado por una *rotación irracional.* . Sea θ un ángulo para el cual $n\theta$ (con $n \in \mathbb{Z}$) nunca es múltiplo de 2π, es decir para el que no existen $p, q \in \mathbb{Z}$ tales que $\theta = (p/q)\,2\pi$ (por ejemplo, $\theta = \sqrt{2}\pi$, o $\theta = 1$ pues π como número ya es irracional)[2]. Entonces el conjunto $\langle \mathsf{R}_\theta \rangle := \{\mathsf{R}_{n\theta} \mid n \in \mathbb{Z}\}$ es un grupo de rotaciones; para $n = 0$ tenemos la identidad y la composición de dos claramente está en él. En este caso, si tomamos un punto x distinto del origen (en nuestra notación anterior, $\mathsf{F}_0 = \{x\}$) las copias $\mathsf{R}_{n\theta}\,(x)$ van poblando poco a poco (al crecer n en ambos sentidos) el círculo $\mathcal{C}$ con centro en 0 y que contiene a x, pues nunca caen sobre un punto ya "pintado", pero nunca llegan a llenar todo el círculo; lo "pueblan", lo "atascan", son "densos en él" pero no lo llenan. Si definimos $\mathsf{F} = \{\mathsf{R}_{n\theta}\,(x) \mid n \in \mathbb{Z}\}$ entonces $\langle \mathsf{R}_{n\theta} \rangle \subset \mathbf{Sim}\,(\mathsf{F})$ (no pusimos la igualdad pues aparecen reflexiones como simetrías). Sin embargo, esta "figura" no se le aparecería nunca a un artista (o vil ser humano) pintando en un papel con lápiz o plumón de cierto grosor. Para eliminar de nuestras consideraciones este tipo de figuras raras o abstractas, que sólo se le aparecen a un matemático quisquilloso, tendríamos que usar conceptos de "topología de conjuntos" que expresen nuestras nociones intuitivas de "continuidad".

La versión topológica del Teorema de Leonardo consiste en poner restricciones topológicas a las figuras (con "cerradas y acotadas" basta) para concluir que su grupo

[2]Un número es *irracional* si no se puede expresar como una fracción p/q con p y q enteros. Por ejemplo $\sqrt{2}$ o π; aunque sea difícil demostrarlo.

de simetrías es $\mathbf{O}(2)$ o bien finito. Intuitivamente, lo de "cerrado" implica que si sus puntos se acumulan cerca de otro, ese también está, de tal manera que los "huecos" de nuestro último ejemplo se llenarían y las simetrías crecerían a todas las rotaciones y las reflexiones ($\mathbf{O}(2)$). Pero dejemos ya de lado esta motivación para el estudio de la topología de conjuntos (cómo controlar particularidades sobre la "continuidad" de subconjuntos de $\mathbb{R}^2$) y entremos de lleno en lo que sí estamos en posición de hacer: poner restricciones sobre el grupo en vez de sobre la figura; usar técnicas de geometría discreta. Demostraremos entonces la versión "grupera" del Teorema de Leonardo.

Versión grupera

Teorema 3.26 (Leonardo) *Si* G *es un grupo finito de isometrías del plano entonces* G *es diédrico o cíclico.*

La demostración consta de cuatro pasos muy bien diferenciados y con ideas de distinta índole en cada uno.

Paso 1 (El centro de simetría). Como G es un grupo finito, sea $n \geq 1$ su cardinalidad. Entonces podemos enumerar sus elementos como

$$G = \{g_0, g_1, \ldots, g_{n-1}\},$$

donde no importa el orden, salvo que podemos distinguir a g_0 como la identidad, que sabemos que está en G; y digamos de una vez que a $g_0 = \mathrm{id}_{\mathbb{R}^2}$ también lo llamamos g_n para facilitar la notación.

Vamos a demostrar que existe un punto $\mathbf{c} \in \mathbb{R}^2$ *invariante bajo* G, es decir, tal que $g_i(\mathbf{c}) = \mathbf{c}$ para toda $i = 1, \ldots, n$, que es naturalmente el *centro de simetría* del grupo. Encontrarlo es fácil. Tomamos cualquier punto $\mathbf{x}$ de $\mathbb{R}^2$, y $\mathbf{c}$ es el baricentro de su *órbita* $G\mathbf{x} := \{g_1(\mathbf{x}), \ldots, g_n(\mathbf{x})\}$. Es decir, sea

$$\mathbf{c} := \sum_{i=1}^{n} \left(\frac{1}{n}\right) g_i(\mathbf{x}).$$

Dibujo (Estamos generalizando la noción de baricentro de un triángulo, 3 puntos, a conjuntos arbitrarios, pero finitos, de puntos.) Para ver que $\mathbf{c}$ es invariante bajo G, usamos el hecho de que las isometrías son funciones afines y que éstas preservan combinaciones afines (Teorema 3.22), pues definimos $\mathbf{c}$ como una combinación tal ($\sum_{i=1}^{n}(1/n) = 1$). Se tiene entonces que para cualquier $g \in G$ (podemos eliminar su subíndice para aliviar notación):

$$g(\mathbf{c}) = \sum_{i=1}^{n} \left(\frac{1}{n}\right) g(g_i(\mathbf{x})) = \sum_{i=1}^{n} \left(\frac{1}{n}\right) (g \circ g_i)(\mathbf{x}). \tag{3.17}$$

Afirmamos que esta última sumatoria es justo la original (que definió a $\mathbf{c}$) salvo que el orden de los sumandos está permutado. Esto se sigue de que "componer con g" es una biyección del grupo en sí mismo (una permutación de los elementos del grupo).

Más formalmente, definamos la función *multiplicación* por g como

$$\mu_g : G \to G$$
$$\mu_g(g_i) = g \circ g_i.$$

Que está bien definida (que siempre caemos en elementos de G) se sigue de que G es cerrado bajo composición; y para ver que es biyección basta exhibir su inversa: $\mu_{g^{-1}}$, multiplicar por g^{-1} (de nuevo bien definida pues $g^{-1} \in G$). Claramente se tiene que $\mu_{g^{-1}} \circ \mu_g = \mathrm{id}_G$ pues

$$\left(\mu_{g^{-1}} \circ \mu_g\right)(g_i) = \mu_{g^{-1}}\left(\mu_g(g_i)\right) = \mu_{g^{-1}}(g \circ g_i) = g^{-1} \circ g \circ g_i = g_i$$

y análogamente, $\mu_g \circ \mu_{g^{-1}} = \mathrm{id}_G$.

Así que la combinación afín que da $g(\mathbf{c})$ (3.17) es un simple reordenamiento de la que da $\mathbf{c}$ y por lo tanto $g(\mathbf{c}) = \mathbf{c}$ para toda $g \in G$. Hemos demostrado que G tiene centro de simetría y, en particular, que no contiene ni traslaciones ni pasos pues estas transformaciones no tienen puntos fijos.

Paso 2 (Vámonos al origen). Ahora veremos que podemos suponer que el centro de simetría de G es el origen, y simplificar así nuestras consideraciones subsecuentes, usando nuestro viejo truco de conjugar con la translación de $\mathbf{c}$ al origen. Más formalmente, sea

$$G_0 = \tau_{-\mathbf{c}} \circ G \circ \tau_{\mathbf{c}} := \{\tau_{-\mathbf{c}} \circ g \circ \tau_{\mathbf{c}} \mid g \in G\},$$

(el conjunto de los conjugados de todos los elementos de G). Nótese que si $g \in G$ entonces $(\tau_{-\mathbf{c}} \circ g \circ \tau_{\mathbf{c}})(\mathbf{0}) = \tau_{-\mathbf{c}}(g(\tau_{\mathbf{c}}(\mathbf{0}))) = \tau_{-\mathbf{c}}(g(\mathbf{c})) = \tau_{-\mathbf{c}}(\mathbf{c}) = \mathbf{0}$ así que $\mathbf{0}$ es el centro de simetría de G_0. Es muy fácil ver que G_0 también es un grupo y que es *isomorfo* al grupo original G (si G fuera el grupo de simetrías de una figura F, G_0 es el grupo de simetrías de $\tau_{-\mathbf{c}}(F)$ —F trasladada al origen— y ambas figuras claramente tienen la "misma" simetría), pues la conjugación respeta composición

$$\tau_{-\mathbf{c}} \circ (h \circ g) \circ \tau_{\mathbf{c}} = \tau_{-\mathbf{c}} \circ h \circ (\tau_{\mathbf{c}} \circ \tau_{-\mathbf{c}}) \circ g \circ \tau_{\mathbf{c}} = (\tau_{-\mathbf{c}} \circ h \circ \tau_{\mathbf{c}}) \circ (\tau_{-\mathbf{c}} \circ g \circ \tau_{\mathbf{c}})$$

e inversos $((\tau_{-\mathbf{c}} \circ g \circ \tau_{\mathbf{c}})^{-1} = \tau_{\mathbf{c}}^{-1} \circ g^{-1} \circ \tau_{-\mathbf{c}}^{-1} = \tau_{-\mathbf{c}} \circ g^{-1} \circ \tau_{\mathbf{c}})$. De hecho, la función inversa de la conjugación es la conjugación por la inversa, $G = \tau_{\mathbf{c}} \circ G_0 \circ \tau_{-\mathbf{c}}$. Así que si entendemos G_0 (con centro de simetrías $\mathbf{0}$) entenderemos el grupo original G. Simplificando nuestra notación, suponemos en adelante que G tiene como centro de simetría el origen, que todos sus elementos dejan fijo el $\mathbf{0}$ y, por lo tanto, que es subgrupo de las transformaciones ortogonales. Es decir, en adelante $G \subset \mathbf{O}(2)$.

Paso 3 (Cuando G preserva orientación). Supongamos que todos los elementos de $G \subset \mathbf{O}(2)$ preservan orientación. Entonces todos los elementos de G son rotaciones con un cierto ángulo θ y tenemos una nueva enumeración natural (no necesariamente la que habíamos usado, salvo la primera) usando sus ángulos:

$$\begin{aligned} G &= \left\{R_{\theta_0}, R_{\theta_1}, R_{\theta_2}, \ldots, R_{\theta_{n-1}}\right\}, \\ \text{donde } \ 0 &= \theta_0 < \theta_1 < \theta_2 < \cdots < \theta_{n-1} < 2\pi. \end{aligned}$$

Queremos demostrar que $\theta_1 = 2\pi/n$ y que $\theta_i = i\,\theta_1$, lo cual es justo decir que $G = \mathbf{C}_n$, que G es el grupo cíclico de orden n generado por la rotación $R_{2\pi/n}$. Puesto que θ_1 es el más chico de estos ángulos (positivos) y todos sus múltiplos $2\theta_1, 3\theta_1, \ldots$ están en el conjunto de ángulos de G (pues se obtienen al componer R_{θ_1} consigo misma), entonces la primera vez que $k\theta_1$ exceda a 2π tiene que ser justo por θ_1 y después de haber pasado por todos los θ_i; es decir, $\theta_1 = 2\pi/n$.

Paso 4 (Jaque mate). Supongamos, por último, que algún elemento de G, digamos que f_0, no preserva orientación. Consideremos el subgrupo de G que sí preserva orientación, y llamémosle G^+; es decir, sea

$$G^+ := \{g \in G \mid \det(g) = 1\}.$$

Claramente G^+ es un grupo (contiene a la identidad y tanto el producto como los inversos de transformaciones que preservan orientación preservan orientación) que, por el paso anterior, es un grupo cíclico $\mathbf{C}_m$ para alguna $m < n$ (como hasta ahora, n es el *orden* del grupo G, es decir, $n = \sharp G$).

Primero veamos que $2m = n$. En el Paso 1 definimos la biyección de G "multiplicar por _"; veamos qué hace μ_{f_0} (multiplicar por nuestro elemento dado que no preserva orientación). Al componer con una transformación que invierte orientación (f_0), a una que la preserva ($g \in G^+$) nos da una que la invierte ($\mu_{f_0}(g) \notin G^+$) y a una que la invierte ($g \notin G^+$), resulta que preserva ($\mu_{f_0}(g) \in G^+$); podemos resumir este trabalenguas como

$$g \in G^+ \quad \Leftrightarrow \quad \mu_{f_0}(g) \notin G^+.$$

Esto quiere decir que μ_{f_0} es una biyección entre G^+ y su complemento, G^- digamos —y vale la notación aunque no sea grupo—, de tal manera que ambos tienen el mismo número de elementos, m habíamos dicho, y entonces $n = 2m$. Por el paso anterior,
Dibujo G^+ es cíclico de orden m y está generado por la rotación $R_{2\pi/m}$.

Por otro lado, como $G \subset \mathbf{O}(2)$, f_0 tiene que ser una reflexión en una línea ℓ_0 que pasa por el origen. En el Ejercicio 3.82 se pidió demostrar analíticamente que $f_0 \circ R_\theta$ es la reflexión en la línea $R_{-\theta/2}(\ell_0)$. Geométricamente, si aplicamos $f_0 \circ R_\theta$ a está última línea, R_θ la gira a la que está a un ángulo $\theta/2$ más allá de ℓ_0 y luego f_0 la regresa a su lugar, así que es cierto lo que se pidió (aunque no hayan hecho su tarea). De tal manera que

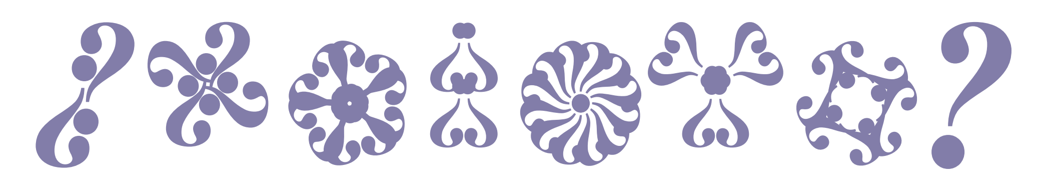
$$f_1 := \mu_{f_0}\left(R_{2\pi/m}\right) = f_0 \circ R_{2\pi/m} \in G^-$$

es la reflexión en la recta ℓ_1 cuyo ángulo hacia ℓ_0 es π/m. Por nuestra definición de grupo diédrico, las reflexiones f_0 y f_1 generan a $\mathbf{D}_m$; pero también generan a G y al verlo concluiremos. A saber, $R_{2\pi/m}$ se obtiene como $f_0 \circ f_1$, pues $f_0 = f_0^{-1}$, así que generan a G^+ (generado por $R_{2\pi/m}$) y luego cualquier elemento de G^- se obtiene como $\mu_{f_0}(g) = f_0 \circ g$ para algún $g \in G^+$. Así que $G = \mathbf{D}_m$ y "colorín colorado, el Teorema de Leonardo ha quedado demostrado". □

Vale la pena hacer notar de la demostración anterior que, aunque el grupo cíclico quede totalmente definido al fijar su centro de simetría (la **c** que encontramos en el Paso 1), o bien que esté conjuntistamente determinado como subgrupo de $\mathbf{O}(2)$, no sucede lo mismo para el diédrico, o **los** diédricos para ser más claros. Nótese que nunca usamos el ángulo de la recta ℓ (el espejo básico) en la que se basó (valga la redundancia) el final de la demostración. Para cualquier recta ℓ que pasa por el origen funciona la demostración y tenemos una instancia, conjuntistamente distinta, del diédrico. Si queremos **el diédrico** con ℓ igual al eje de las x, tendremos que conjugar, como en el Paso 2, pero ahora con la rotación que lleva ℓ a la recta canónica.

Quizá quede más claro a qué nos referimos con eso de "instancias" de un cierto grupo si enunciamos el Teorema de Leonardo como "*un grupo finito de isometrías es conjugado del cíclico* $\langle R_{2\pi/n} \rangle$ *o del diédrico* $\langle E_0, E_{\pi/n} \rangle$" que fue, junto con la observación precedente, lo que realmente demostramos.

EJERCICIO 3.104 Identifica el grupo de simetrías de las siguientes figuras y, a ojo de buen cubero, píntales su centro de simetría. ¿Qué algunas no tienen tal? ¿Dónde está un error formal (no esencial) de la demostración anterior? Corrígelo.

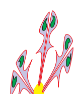


Leonardo afín

El ojo humano, y la naturaleza en general, identifica simetrías aunque, estrictamente hablando, no las haya; vemos el bulto y no nos preocupamos por los milímetros; el ejercicio anterior tiene sentido aunque es seguro que en el dibujo, o en la impresión, se cometieron errores; el cuerpo humano tiene simetría $\mathbf{D}_1$. E igualmente existe la simetría más allá del rígido grupo de isometrías.

Supongamos que una figura F tiene grupo de simetrías G (será $\mathbf{D}_6$ en la figura que viene). Si f es una transformación afín que no es isometría, entonces $f(F)$ (ahora sí: la figura) ya no tiene, en general y estrictamente hablando, las mismas simetrías. Sin embargo, es evidente que sí las tiene: las vemos, las detectamos: tiene simetría afín. Podemos definir el *grupo de simetría afín* de (cualquier figura) F (luego regresamos a nuestro ejemplo, $f(F)$) como

$$\mathbf{Sim}_{\mathbf{Af}}(F) = \{g \in \mathbf{Af}(2) \mid g(F) = F\}$$



y la demostración de que es un subgrupo de $\mathbf{Af}(2)$ es, *verbatim*,[3] la de que $\mathbf{Sim}(\mathsf{F})$ es subgrupo de $\mathbf{Iso}(2)$.

En nuestro ejemplo ($\mathsf{f}(\mathsf{F})$, el dibujo de al lado), $\mathbf{Sim}_{\mathbf{Af}}(\mathsf{f}(\mathsf{F}))$ será justo el conjugado por f^{-1} del grupo $\mathsf{G} = \mathbf{Sim}(\mathsf{F})$; que, puesto que f no es isometría, simplemente es afín, sólo es un subgrupo de las transformaciones afines. Veamos. Si $\mathsf{g} \in \mathsf{G}$ entonces

$$\begin{aligned} \left(\mathsf{f} \circ \mathsf{g} \circ \mathsf{f}^{-1}\right)(\mathsf{f}(\mathsf{F})) &= (\mathsf{f} \circ \mathsf{g})\left(\mathsf{f}^{-1}(\mathsf{f}(\mathsf{F}))\right) \\ &= (\mathsf{f} \circ \mathsf{g})(\mathsf{F}) = \mathsf{f}(\mathsf{g}(\mathsf{F})) = \mathsf{f}(\mathsf{F}) \end{aligned}$$

implica que $\left(\mathsf{f} \circ \mathsf{g} \circ \mathsf{f}^{-1}\right) \in \mathbf{Sim}_{\mathbf{Af}}(\mathsf{f}(\mathsf{F}))$. Hemos demostrado que

$$\mathsf{f} \circ \mathbf{Sim}(\mathsf{F}) \circ \mathsf{f}^{-1} \subset \mathbf{Sim}_{\mathbf{Af}}(\mathsf{f}(\mathsf{F})),$$

donde estamos extendiendo naturalmente nuestra definición de grupo conjugado (dada en el Paso 2 del Teorema de Leonardo) al contexto más amplio de cualquier grupo (en nuestro caso a $\mathbf{Af}(2)$). Intentemos demostrar la otra contención, pues habíamos aventurado justo la igualdad.

dibujo arriba

Si $\mathsf{g} \in \mathbf{Sim}_{\mathbf{Af}}(\mathsf{f}(\mathsf{F}))$, entonces un cálculo equivalente, aunque más fácil, implica que $\left(\mathsf{f}^{-1} \circ \mathsf{g} \circ \mathsf{f}\right)(\mathsf{F}) = \mathsf{F}$. Pero de esto no podemos concluir que $\left(\mathsf{f}^{-1} \circ \mathsf{g} \circ \mathsf{f}\right) \in \mathbf{Sim}(\mathsf{F})$ sino sólo que $\left(\mathsf{f}^{-1} \circ \mathsf{g} \circ \mathsf{f}\right) \in \mathbf{Sim}_{\mathbf{Af}}(\mathsf{F})$, pues no sabemos si $\left(\mathsf{f}^{-1} \circ \mathsf{g} \circ \mathsf{f}\right)$ es isometría o no: en principio, es simplemente afín. Habíamos aventurado la igualdad justa en el caso concreto del dibujo $\mathsf{f}(\mathsf{F})$ que estamos estudiando, porque a simple vista es evidente: el "chuequito" $\mathsf{f}(\mathsf{F})$ tiene únicamente las 12 simetrías afines que corresponden a las 12 simetrías métricas del "redondito" F. Y aunque no lo hayamos podido demostrar en el primer intento, surgen las preguntas naturales de si los grupos de simetrías afines de las figuras acotadas (versión topológica) o si los subgrupos finitos (versión grupera) se comportan como los rígidos (teoremas de Leonardo). Más en concreto, y en base a lo que hemos logrado con isometrías, olvidándonos de las figuras y concentrándonos en los grupos, nos debemos hacer dos preguntas:

Pregunta 1. *Si* G *es un subgrupo finito de* $\mathbf{Af}(2)$, *¿será cierto que es isomorfo a un cíclico o un diédrico?*

Pregunta 2. *Si* G *es un subgrupo finito de* $\mathbf{Af}(2)$, *¿será cierto que es conjugado del cíclico* $\left\langle \mathsf{R}_{2\pi/n} \right\rangle$ *o del diédrico* $\left\langle \mathsf{E}_0, \mathsf{E}_{\pi/n} \right\rangle$ *para alguna* n*?*

Dejamos estas preguntas al lector, pues gran parte del trabajo ya se hizo. Por ejemplo, el Paso 1 en la demostración de Leonardo funciona *verbatim* en el caso afín (de hecho sólo en esa cualidad se basó); y ya que lo mencionamos, ahí está el error que dejamos buscar como ejercicio: para que el centro $\mathbf{c}$ esté bien definido, que no dependa de $\mathbf{x}$, se necesita que el grupo tenga más de dos elementos. Y el Paso 2, irse al origen, es igual, pero acabamos obteniendo que G es subgrupo de $\mathbf{GL}(2)$ en vez de $\mathbf{O}(2)$. Los problemas aparecen en el Paso 3 que aparentemente se basó en

[3] Dícese de cuando es literalmente idéntica, salvo los cambios obvios.

sumas, restas y desigualdades de ángulos como números reales, pero en el fondo son una forma más mundana de trabajar con la estructura de grupo, y el orden cíclico que dan se puede obtener al tomar la órbita ($G\mathbf{x} = \{g(\mathbf{x}) \mid g \in G\}$) de algún punto. Algunos de los ejercicios siguientes pueden ayudar.

EJERCICIO 3.105 Sea $g \in \mathbf{GL}^+(2)$ (donde $\mathbf{GL}^+(2) := \{g \in \mathbf{GL}(2) \mid \det(g) > 0\}$), tal que para algún vector $\mathbf{v} \in \mathbb{R}^2$, $\mathbf{v} \neq \mathbf{0}$, se tiene que $g(\mathbf{v}) = t\mathbf{v}$. Demuestra que si g tiene orden finito, es decir, que si $g^n = \mathrm{id}$ para alguna n, entonces $g = \mathrm{id}$ o bien $g = -\mathrm{id}$. *(Primero demuestra que* $t = \pm 1$ *pensando en la acción de* g *sobre la recta generada por* $\mathbf{v}$, $\langle \mathbf{v} \rangle$*; y luego usa la hipótesis de que preserva orientación: toma algún vector fuera de* $\langle \mathbf{v} \rangle$ *y expresa lo que le hace* g*).*

EJERCICIO 3.106 Demuestra que si $G \subset \mathbf{GL}^+(2)$ es finito, entonces $G\mathbf{v} = \{g(\mathbf{v}) \mid g \in G\}$ tiene la cardinalidad de G para cualquier $\mathbf{v} \neq \mathbf{0}$. Y que además se tiene que si $g(\mathbf{v})$ es paralelo a $h(\mathbf{v})$ para algunos g y h en G, entonces $g = h$ o bien $g = -h$ (*ojo: en este último caso, sus ángulos son opuestos*).

EJERCICIO 3.107 Sea $g \in \mathbf{GL}(2)$, $g \neq \mathrm{id}$; demuestra que su conjunto de puntos fijos, $\left\{\mathbf{x} \in \mathbb{R}^2 \mid g(\mathbf{x}) = \mathbf{x}\right\}$, es una recta que pasa por el origen o sólo el origen (*revisa la Sección 8.1*). Concluye que los puntos fijos de una transformación afín distinta de la identidad son el vacío, o un punto o una recta.

EJERCICIO 3.108 Demuestra que $\mathbf{Sim}_{\mathbf{Af}}\left(\mathbb{S}^1\right) = \mathbf{O}(2) = \mathbf{Sim}\left(\mathbb{S}^1\right)$; es decir, que toda simetría afín del círculo unitario es una isometría.

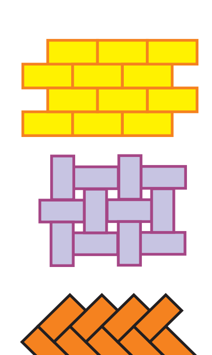
EJERCICIO 3.109 Sea $\mathcal{E}$ la elipse dada por la ecuación

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

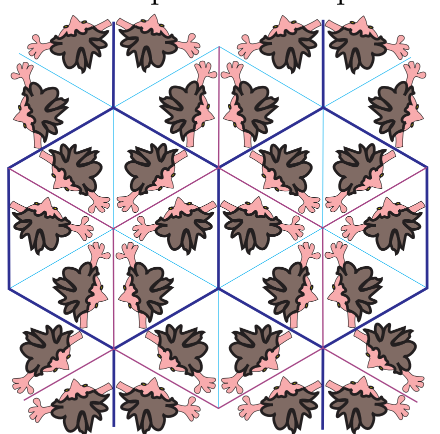
Demuestra que $\mathbf{Sim}_{\mathbf{Af}}(\mathcal{E})$ es conjugado de $\mathbf{O}(2)$ (para esto, encuentra una matriz que mande a $\mathbb{S}^1$ en $\mathcal{E}$).

3.9.2 Grupos discretos y caleidoscópicos

Los teoremas tipo Leonardo hablan de las posibles simetrías de las figuras acotadas. Pero además hay figuras infinitas que tienen simetría, y de ellas, más bien de sus grupos, queremos hablar en esta sección. Las más comunes son los pisos por los que caminamos a diario y las paredes que nos rodean. Aunque no sean infinitos, una pequeña porción sugiere claramente cómo debía continuarse de tal manera que cada región sea "igualita" a cualquier otra. Tienen, o dan la idea, de una "regla" para ir pegando piezas que llamaremos *losetas* y que juntas todas ellas llenan el plano. A tal descomposición del plano en losetas la llamaremos un *mosaico*; cada loseta es copia isométrica de una o varias que forman el *muestrario*; por ejemplo, en el mosaico de en medio un "ladrillo" rectangular y un "huequito" cuadrado forman el muestrario y en los otros dos el muestrario consta de una sola pieza. Con una sola pieza se pueden armar varios mosaicos (ver otros ejemplos, además de los dos al margen, en el segundo ejercicio que viene, el 3.111) usando diferentes "reglas". Y esto de las "reglas" ya lo conocemos: consisten en el fondo de dar un subgrupo de isometrías; que llamaremos el *grupo de simetrías* del mosaico: las isometrías que lo mandan en sí mismo (que dejan a la figura completa —aunque infinita— en su lugar).



Un último y sugestivo ejemplo, antes de entrar a la teoría, surge de los caleidoscopios clásicos que en su versión física, real, consisten de tres espejos rectangulares del mismo tamaño que forman un prisma triangular equilátero, de tal manera que al observar desde un extremo de él, la figura plana al otro lado (que puede ser cambiante) se reproduce para dar la impresión de un plano infinito. Este es otro ejemplo de un mosaico plano. Podemos pensar en una figura plana (*la figura fundamental*) dentro de un triángulo equilátero de líneas espejo: ésta se reproduce por las tres reflexiones y sus composiciones para tapizar todo el plano y generar una *figura caleidoscópica*; como mosaico, el muestrario consiste de una loseta que es el triángulo equilátero con un dibujo.



El grupo de simetrías de este mosaico es el conjunto de todas las posibles composiciones de las tres reflexiones básicas o *generadoras*; y un elemento del grupo lleva a la figura (o loseta) fundamental a una cierta loseta, que no sólo es un triángulo de la descomposición del plano en triángulos equiláteros, sino que la figura (en el dibujo, el "manquito") le da una orientación y una posición específicas. En su versión física, la composición que produce una cierta loseta tiene que ver con los espejos y el orden en que rebota el rayo de luz que emana de la figura real para producir la imagen dada.

Basta de ejemplos (ojéense los ejercicios) y hagamos algo de teoría. Por nuestra experiencia reciente, será más facil definir con precisión a los grupos involucrados en los mosaicos que a los mosaicos mismos (se necesitarían rudimentos de topología). Pero la intuición y los ejemplos bastan. Aunque los grupos de simetría de los mo-

saicos sean infinitos, tienen la propiedad de que localmente son finitos; cada punto del plano está en un número finito de losetas. Esto se reflejará en la primera propiedad de nuestra definición. Y otra cosa importante es que las losetas sean acotadas, que expresaremos como la propiedad dos. Llamaremos *disco* a un círculo (de radio positivo) junto con su interior, es decir a los conjuntos de la forma $D = \left\{ \mathbf{x} \in \mathbb{R}^2 \mid d(\mathbf{x}, \mathbf{c}) \leq r \right\}$ para algún punto $\mathbf{c} \in \mathbb{R}^2$ y algún $r > 0$.

Definición 3.9.1 Un subgrupo G de las isometrías **Iso** (2) es un *grupo cristalográfico* si cumple:

i) Para cualquier disco D se tiene que el conjunto $\{g \in G \mid g(D) \cap D \neq \emptyset\}$ es finito.

ii) Existe un disco D_0 tal que $\bigcup_{g \in G} g(D_0) = \mathbb{R}^2$, es decir, tal que para todo $\mathbf{x} \in \mathbb{R}^2$ se tiene que $\mathbf{x} \in g(D_0)$ para alguna $g \in G$.

La primera propiedad dice que al estudiar un grupo cristalográfico dentro de una región acotada del plano sólo intervendrá un número finito de elementos del grupo; a esto también se le llama que el grupo sea *discreto*. Y la segunda, que al entender una región suficientemente grande, pero acotada, de alguna manera ya acabamos; el grupo se encarga del resto. Un ejemplo importante son los subgrupos generados por dos traslaciones en direcciones independientes, por ejemplo en los vectores canónicos $\boldsymbol{e}_1$ y $\boldsymbol{e}_2$, que da lugar a un subgrupo isomorfo a $\mathbb{Z} \times \mathbb{Z}$ (en el ejemplo son precisamente las traslaciones por vectores en $\mathbb{Z}^2 \subset \mathbb{R}^2$).

El nombre "cristalográfico" viene de que los cristales en la naturaleza se reproducen respetando ciertas simetrías; los químicos, interesados en el asunto, se pusieron a estudiar tales simetrías y demostraron, hacia el final del siglo XIX, que en el plano solamente había 17 posibilidades teóricas para estas reglas de crecimiento. Los matemáticos, ya en el siglo XX, pulieron sus observaciones y la demostración, pero preservaron el nombre dado por los químicos.

La demostración de este teorema involucra dos grandes pasos: la existencia (exhibir los 17 grupos cristalográficos) y la unicidad (demostrar que no hay más). Pero resulta que los artistas y artesanos árabes —cuya expresión gráfica se centró mucho en los mosaicos, la simetría y la ornamentación pues en cierta interpretación del Corán se prohibía representar la figura humana— ya habían hecho (¡y de qué manera!) la mitad del trabajo: en la Alhambra, hay ejemplos de mosaicos con los 17 grupos. Así que la demostración de la existencia está en la Alhambra, concluida en el siglo XVI; y si hay un Teorema de Leonardo, éste —digo yo— debía ser el de la Alhambra.

No tenemos tiempo para demostrarlo aquí. Pero sí podemos ejemplificar su demostración con uno mucho más modesto. Diremos que un grupo cristalográfico G es *caleidoscópico* si además cumple con que tiene suficientes reflexiones como para generarlo, es decir, si cada elemento de G puede escribirse como composición de reflexiones en G. Un ejemplo es el que describimos arriba, que estaba generado por tres reflexiones y que denotaremos como $K_{3,3,3}$ pues las líneas de sus espejos se encuentran en los ángulos $\pi/3, \pi/3, \pi/3$. Dibujo

Otros dos grupos caleidoscópicos surgen del juego de escuadras. Con tres lineas espejo en los lados de un triángulo de ángulos $\pi/6, \pi/3, \pi/2$ (la escuadra "$30, 60, 90$") se genera el grupo $K_{6,3,2}$; y con tres espejos en ángulos $\pi/4, \pi/4, \pi/2$ (la escuadra "45") el $K_{4,4,2}$. El último, que llamamos $K_{4\times4}$, es el generado por espejos en los cuatro lados de un rectángulo. Y, salvo ver que efectivamente son cristalográficos en cuyos detalles no entraremos (construyánse modelos con espejos reales), esto concluye la parte de existencia del siguiente teorema

Teorema 3.27 *Hay cuatro grupos caleidoscópicos, que son* $K_{6,3,2}$, $K_{4,4,2}$, $K_{3,3,3}$ *y* $K_{4\times4}$.

La demostración de la unicidad se basa en un lema general sobre los centros de simetría. Dado un grupo cristalográfico G (aunque de hecho sólo usaremos que es discreto, la propiedad **(i)**) tenemos que para cualquier punto $x \in \mathbb{R}^2$ hay un subgrupo de G, llamado el *estabilizador* de x y denotado con $\mathbf{St}_G(x)$, que consta de los elementos de G que no mueven a x, que lo "estabilizan", es decir,

$$\mathbf{St}_G(x) := \{g \in G \mid g(x) = x\}.$$

(Es fácil demostrar que efectivamente es un subgrupo de G.) Si tomamos cualquier disco D centrado en x, todos los elementos de su estabilizador mandan a D precisamente en D, pues fijan su centro y son isometrías. Por la propiedad **(i)** los que mandan a D cerca (lo suficiente para intersectarlo) ya son un número finito, entonces $\mathbf{St}_G(x)$ es finito para cualquier $x \in \mathbb{R}^2$. El Teorema de Leonardo nos da entonces el siguiente lema.

Lema 3.28 *Si* G *es un grupo discreto, entonces sus subgrupos estabilizadores no triviales son cíclicos o diédricos.* □

Ahora sí, enfoquemos nuestras baterías contra el teorema de los grupos caleidoscópicos. Como el grupo G tiene suficientes reflexiones para generarlo, nos podemos fijar en el conjunto de sus líneas espejo. Es decir, para cada reflexión en G, pintamos de negro su línea espejo; llamemos $\mathcal{L}$ a este conjunto de lineas. Primero vamos a demostrar que podemos encontrar un triángulo o un rectángulo formado por líneas negras (en $\mathcal{L}$) tal que ninguna línea de $\mathcal{L}$ pasa por su interior y lo llamaremos la *región fundamental.*

Dibujo

Si todas las líneas de $\mathcal{L}$ son paralelas, los elementos de G mueven cualquier disco en la dirección perpendicular y entonces cubren a lo más una franja del tamaño de su diámetro en esa dirección y no existiría el disco D_0 de la condición **(ii)**. Hay, por tanto, al menos dos pendientes en $\mathcal{L}$.

Si son exactamente dos las pendientes en $\mathcal{L}$, éstas tienen que ser perpendiculares, pues con reflexiones en dos rectas no ortogonales se generan reflexiones en líneas con nuevas direcciones. En este caso, hay al menos dos rectas de cada tipo, pues si hay sólo una de un tipo, con cualquier disco se cubre, de nuevo, a lo más una franja.

Dibujo

Existe entonces un rectángulo en $\mathcal{L}$.

Si hay rectas en tres direcciones tenemos un triángulo. Llamemos F' al rectángulo o triángulo que hemos encontrado. Puesto que las reflexiones en G que mandan a F' cerca son finitas (la condición **(i)** aplicada a un disco que lo contenga), hay, a lo más, un número finito de rectas en $\mathcal{L}$ que pasan por el interior de F'. Recortando con tijeras en todas ellas, un pedacito (cualquiera) de lo que queda es la región fundamental F que buscábamos, y ya no pasan lineas negras por su interior.

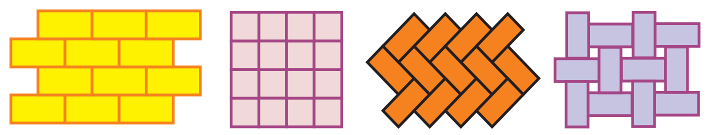
Veamos ahora cómo puede ser la región fundamental F. En un vértice, $\mathbf{v}$ digamos, se intersectan dos líneas de $\mathcal{L}$ de tal manera que en su ángulo definido por F (como ángulo interior) no hay ninguna otra línea de $\mathcal{L}$. Como las reflexiones correspondientes a estas dos líneas son elementos del grupo estabilizador de $\mathbf{v}$, $\mathrm{St}_G(\mathbf{v})$, y éste es diédrico o cíclico según el Lema 3.28, entonces es diédrico (pues tiene reflexiones) y el ángulo interior de F en $\mathbf{v}$ es de la forma π/n. Por lo tanto, F tiene que ser un rectángulo (con cuatro ángulos $\pi/2$), o bien un triángulo con ángulos $\pi/p, \pi/q, \pi/r$ donde podemos suponer que $p \leq q \leq r$. En el segundo caso se tiene entonces que p, q, r cumplen

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1,$$

pues los ángulos de un triángulo suman π. Para $p = 2$, tenemos las soluciones $q = 3$, $r = 6$ y $q = r = 4$; para $p = 3$ tenemos la solución $q = r = 3$ y para $p \geq 4$ ya no hay soluciones.

Entonces hemos demostrado que los lados de F generan uno de los grupos $K_{4\times4}$, $K_{6,3,2}$, $K_{4,4,2}$ o $K_{3,3,3}$ (respectivamente a su orden de aparición) y por lo tanto que uno de estos grupos, el correspondiente a F, digamos que G', está contenido en G. Dejamos demostrar que $G' = G$ al lector intrigado; tendrá que usar sustancialmente que las reflexiones generan a G.

EJERCICIO 3.110 Identifica las *líneas de simetría* (espejos de reflexiones y líneas de pasos en el grupo de simetrías) y los *centros de simetría* (centros de subgrupos finitos), de los siguientes mosaicos comunes, que, por supuesto, suponemos infinitos.



EJERCICIO 3.111 Supón que en los siguientes mosaicos la loseta oscura está justo en el cuadrado unitario. Da generadores para sus grupos de simetría.

Rearmar figura co mosaiquito asimétrico

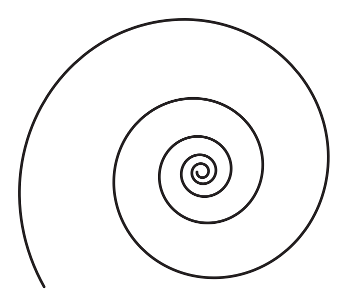
[3] Para leer más sobre simetría y mosaicos ver [15], [8] y [11].

3.9.3 Fractales afinmente autosimilares

Veremos por último algunas figuras que tienen cierta "simetría" aunque no dentro del grupo de isometrías.

Espirales

Consideremos una espiral infinita. Claramente tiene autosemejanzas: siente uno que "gira y crece proporcionalmente". En efecto, el grupo "espiral"



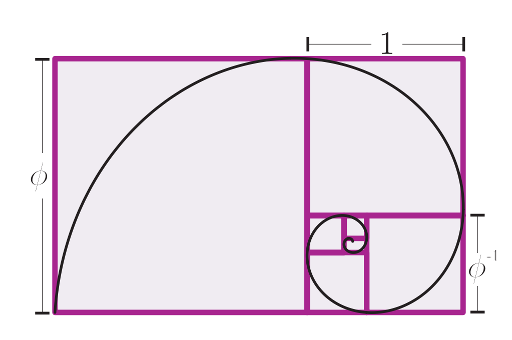
$$\mathsf{E}_c = \left\{ c^t \, R_{t\pi} \mid t \in \mathbb{R} \right\},$$

donde $c > 0$, es un subgrupo de las semejanzas y es isomorfo al grupo de los reales con la suma para cualquier $c \neq 1$, pues

$$\left(c^t \, R_{t\pi}\right) \circ \left(c^r \, R_{r\pi}\right) = c^{(t+r)} \, R_{(t+r)\pi}.$$

Para cualquier punto $\mathbf{x} \neq \mathbf{0}$, su órbita $\mathsf{E}_c(\mathbf{x}) = \{f(\mathbf{x}) \mid f \in \mathsf{E}_c\}$ es una espiral *invariante* bajo el grupo, es decir, su imagen bajo cualquier elemento del grupo es ella misma. Los distintos valores de c dan qué tan rápido crece la espiral y en qué orientación.

Especialmente bonita, y que además se da en la naturaleza, es la *espiral áurea* que crece la proporción áurea ϕ en un cuarto de vuelta y entonces en nuestra fórmula anterior corresponde a $c = \phi^2$. La *proporción áurea*, descubierta y ampliamente usada por los griegos que también la llamaban "proporción divina", es la proporción de los lados de un rectángulo tal que si se le extrae un cuadrado sobre el lado chico nos queda un rectángulo semejante al original. Nótese que en la figura clásica de al lado, considerando que en ella está también la sucesión infinita de cuadrados a la que alude, tiene como grupo de simetrías afines o de semejanzas al generado por la transformación líneal



$$\begin{pmatrix} 0 & -\phi \\ \phi & 0 \end{pmatrix}$$

(poniendo al origen en el centro de la espiral), que es un subgrupo isomorfo a $\mathbb{Z}$ dentro de E_{ϕ^2}.

EJERCICIO 3.112 Demuestra, usando la definición que acabamos de dar, que la proporción áurea cumple con la ecuación

$$x^2 = x + 1$$

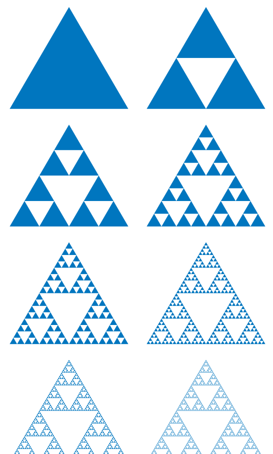
y por tanto que $\phi = \left(1 + \sqrt{5}\right)/2 \simeq 1.61803$.

EJERCICIO 3.113 Si en la figura anterior el cuadrado superior derecho es el cuadrado unitario, ¿dónde está el centro de la espiral?

Fractales autosimilares

Si tomamos una espiral desde su centro hasta algún punto dado, entonces al hacer un *zoom* hacia el centro, sin importar de qué tanto aumento, volvemos a encontrar una copia de la figura original. Veremos ahora cómo definir figuras que cumplen esto pero alrededor de cualquiera de sus puntos.

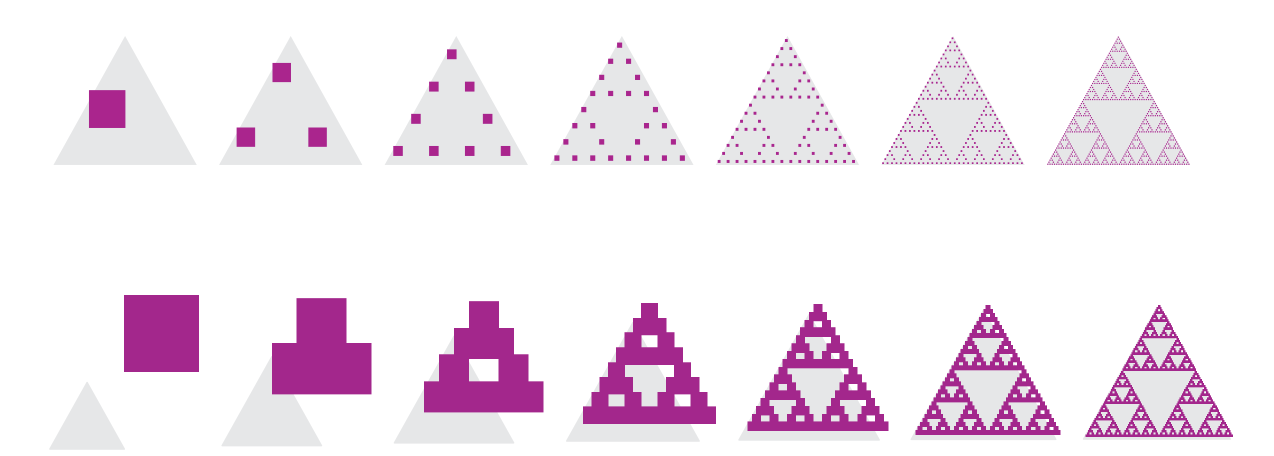
Consideremos primero el ejemplo clásico conocido como el Triángulo de Sierpinsky. Sea Z_0 un triángulo equilátero y sean f_1, f_2 y f_3 las contracciones de factor $1/2$ centradas en sus tres vértices. Tomamos Z_1 como la unión de las tres imágenes de Z_0 bajo estas tres homotecias; y en general, definimos Z_n inductivamente como



$$Z_{n+1} = f_1(Z_n) \cup f_2(Z_n) \cup f_3(Z_n).$$

En la figura vemos hasta Z_7. El triángulo de Sierpinsky Z es el límite de los Z_n, o bien, como $Z_{n+1} \subset Z_n$, también es la intersección de todos ellos, así que, por ejemplo los tres vértices originales están en Z así como todos los vértices de todos los triángulitos (3^n) en cada paso. La propiedad importante que cumple Z es que si magnificamos pequeñas porciones en cualquier lugar vemos copias del original, y se dice que es un *fractal autosimilar*; en este caso, podemos añadir *autosemejante* pues con *zooms* basta.

La clave es que las tres transformaciones afines que consideramos contraen en todas las direcciones y cumplen entonces que independientemente de con qué conjunto acotado empecemos (en vez de Z_0) el límite es el mismo (aunque ya no se obtenga por intersección). No lo demostraremos aquí, pero ponemos dos dibujos del principio de la sucesión correspondiente a otros dos conjuntos, uno un cuadradito dentro del triángulo Z_0 (que se mantiene en gris al fondo) y el otro un cuadrado grande fuera de él. Después de unos cuantos pasos, las tres figuras ya son casi indistinguibles. Así que Z está realmente asociado a las tres semejanzas (reductoras) f_1, f_2 y f_3, y no depende del conjunto acotado con el que empezamos.



Si en vez de semejanzas tomamos transformaciones afines que contraigan en todas las direcciones, es decir, que cumplan la desigualdad

$$d(f(\mathbf{x}), f(\mathbf{y})) < d(\mathbf{x}, \mathbf{y}) \qquad \text{para toda } \mathbf{x}, \mathbf{y} \in \mathbb{R}^2 \text{ con } \mathbf{x} \neq \mathbf{y}, \tag{3.18}$$

se tiene que la figura límite queda compacta y bien definida (teorema que no demostraremos). Así se obtienen fractales afinmente autosimilares.

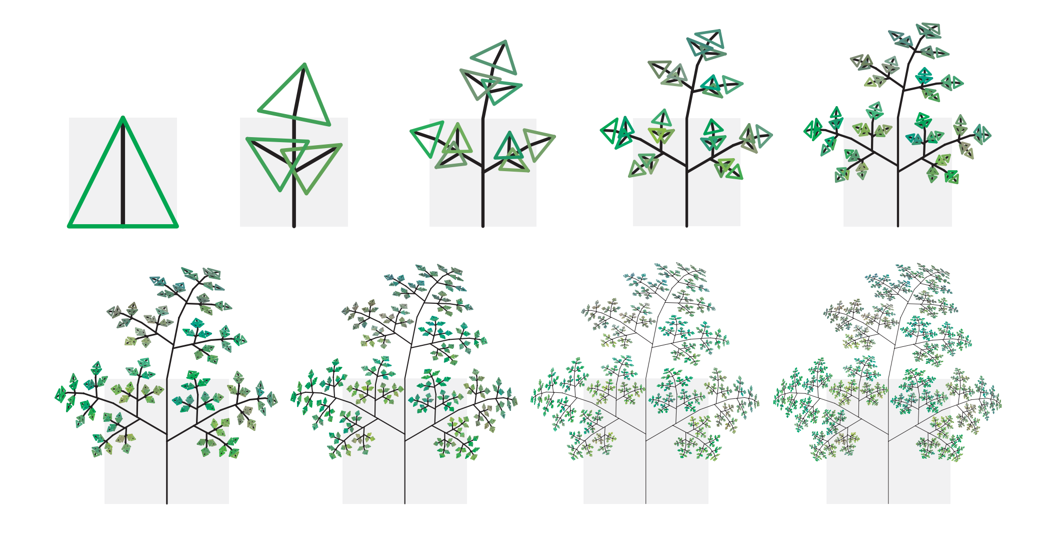
Otro ejemplo interesante es el helecho de la figura siguiente. Está asociado a las tres transformaciones afines g_1, g_2 y g_3 que mandan al triángulo F_0 del primer cuadro en los tres triángulos del segundo respectivamente (en gris se mantiene el cuadrado unitario para mostrar proporciones y de los segmentos hablaremos en breve). Y de nuevo se define inductiva o recursivamente

$$F_{n+1} = g_1(F_n) \cup g_2(F_n) \cup g_3(F_n)$$

para obtener el "follaje" F como el límite de los conjuntos F_n (en la figura aparece hasta F_8); otra vez no depende del conjunto inicial F_0 sino unicamente de las tres transformaciones. (Puesto que reducen área, las tres transformaciones tienen determinante con valor absoluto menor que 1 y una de ellas lo tiene negativo: ¿puedes decir cuál? se aprecia hasta el tercer cuadro.) Obsérvese que los tallos del helecho se obtienen de una manera sutilmente diferente. Si T_0 es el segmento vertical del primer cuadro, entonces a la siguiente iteración también se le añade lo que ya llevabamos —crece, por decirlo botánicamente, sobre lo ya construido—; esto es, definimos

$$T_{n+1} = T_n \cup g_1(T_n) \cup g_2(T_n) \cup g_3(T_n)$$

y entonces el resultado final T (el límite) sí depende de con qué empezamos (y con ese pequeño intervalo se logra el efecto, de "tallo", deseado).



Llamemos $H = T \cup F$ al helecho. Las tres funciones generadoras g_1, g_2, g_3 lo mandan en sus tres ramas principales, las composiciones de un par de ellas lo mandan en sus 9 "ramas secundarias", y así sucesivamente. Si consideramos todas las posibles composiciones finitas de g_1, g_2, g_3, junto con la identidad se obtiene un *semigrupo* (algunos autores lo llaman *grupoide*) que cumple con dos de los axiomas de grupo. A saber, tiene a la identidad y es cerrado bajo composición, pero no tiene inversos; de alguna manera, es el análogo de los grupos de simetría para este tipo de fractales. Y cualquier elemento g de este semigrupo cumple que $g(H) \subset H$; manda al helecho completo en una ramita por ahí que es afinmente equivalente a H.

Como ya mencionamos, es importante que las tres transformaciones afines que usamos para definir el helecho sean contracciones; en particular, entonces F tampoco depende de con qué figura acotada hayamos empezado. Esbozaremos por último, algunas de las ideas que indican el porqué de este hecho.

Primero, para una transformación afín $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, que sea una contracción (que cumpla (3.18)), implica algo más fuerte que llamaremos *contracción fuerte* que es que existe una constante c, $0 < c < 1$ para la cual

$$d(f(\mathbf{x}), f(\mathbf{y})) \leq c\, d(\mathbf{x}, \mathbf{y}) \qquad \text{para toda } \mathbf{x}, \mathbf{y} \in \mathbb{R}^2 \text{ con } \mathbf{x} \neq \mathbf{y}. \tag{3.19}$$

Pues la parte lineal es la que determina las distancias porque

$$|f(\mathbf{x}) - f(\mathbf{y})| = |(A\mathbf{x} + \mathbf{b}) - (A\mathbf{y} + \mathbf{b})| = |A\mathbf{x} - A\mathbf{y}|;$$

y entonces basta ver qué le hace A al círculo unitario. Como quedará claro en el capítulo siguiente, A lo manda en una elipse centrada en el origen y si se cumple (3.18) para $\mathbf{x} = \mathbf{0}$ entonces esta elipse está contenida en el círculo unitario y su semieje mayor $c < 1$ sirve como la constante mínima que cumple (3.19).

Con (3.19) se puede demostrar en general que una contracción fuerte f tiene un punto fijo $\mathbf{p}$. Para una función afín esto es más fácil (véase el Ejercicio 3.115). Y entonces para cualquier $\mathbf{x}$ se tiene que la sucesión $f^n(\mathbf{x})$ tiende (se acerca indefinidamente) a $\mathbf{p}$, pues si d es la distancia de $\mathbf{x}$ a $\mathbf{p}$, es fácil ver que

$$d(f^n(\mathbf{x}), \mathbf{p}) \leq c^n d \to 0$$

(obsérvese que esto no necesariamente es cierto si sólo se cumple (3.18)). Así que las iteraciones bajo f de cualquier conjunto acotado se van aproximando a $\mathbf{p}$ y lo tienen como límite. Esto implica que el punto fijo de cualquier transformación en los dos semigrupos "contractivos" que hemos considerado va a aparecer en el límite (Z o F), y que también lo harán sus imágenes bajo otros elementos del semigrupo. Pero quizás aparezcan más puntos. Por lo menos, esto medio explica por qué el conjunto inicial no importa para definir el fractal.

Por último, hay que mencionar que esto de los fractales es mucho más amplio. Hay de muchos otros tipos y aún la terminología no está bien establecida pues en las

últimas décadas ha habido mucho trabajo sobre ellos y su sorprendente similitud con la naturaleza.

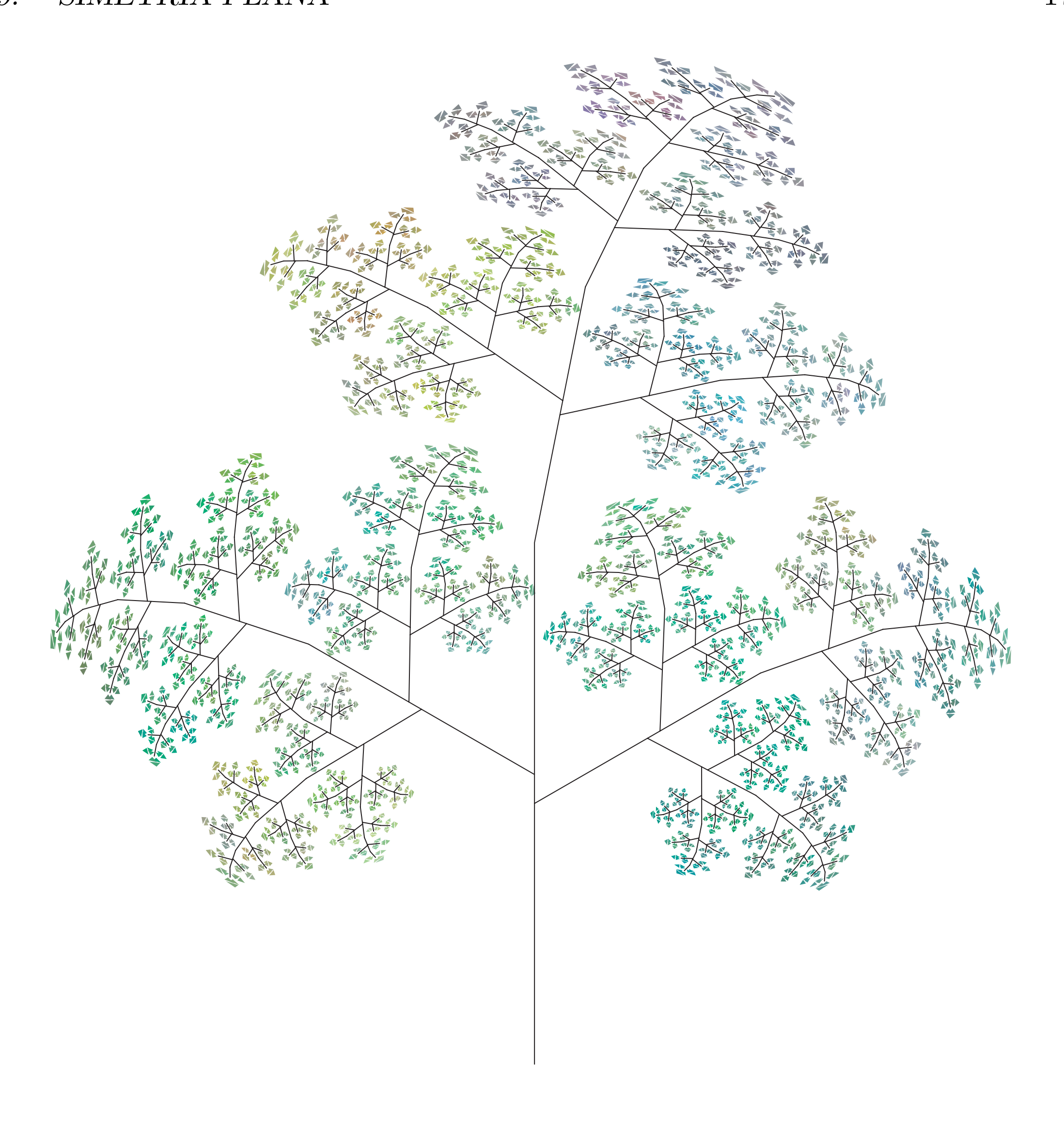
EJERCICIO 3.114 Sea f_1 la transformación afín definitoria del helecho H asociada a su rama principal, es decir, que manda al triángulo F_0 en el más grande y más alto de los tres en F_1, el central, pues. ¿Cuál es su punto fijo? ¿Puedes determinar los puntos fijos de las otras dos? (Uno es fácil y el otro más escurridizo.) ¿De alguna composición doble?

EJERCICIO 3.115 Sea $f(\mathbf{x}) = A\mathbf{x}+\mathbf{b}$ una función afín de $\mathbb{R}^2$ en $\mathbb{R}^2$ que no tiene puntos fijos. Demuestra que entonces existe un vector $\mathbf{v} \neq \mathbf{0}$ para el cual $A\mathbf{v} = \mathbf{v}$. (Usa que la ecuación $f(\mathbf{x}) = \mathbf{x}$ es equivalente a $(A - I)\,\mathbf{x} = -\mathbf{b}$ y en el Teorema 3.21.)

EJERCICIO 3.116 Usando el ejercicio anterior, demuestra que una contracción afín (que cumple (3.18)) tiene un único punto fijo.

---

[3] Como bibliografía general sobre fractales, véanse [18] y [19].



Capítulo 4

Cónicas II (clasificación)

En el Capítulo 2 se introdujeron las cónicas (círculos, elipses, hipérbolas y parábolas) como los lugares geométricos de puntos en el plano euclidiano que cumplen cierta propiedad descrita en términos de distancias. De esa descripción se dedujo que son los conjuntos de puntos cuyas coordenadas cumplen cierta ecuación cuadrática. Una ecuación cuadrática de éstas se puede considerar como un polinomio cuadrático (de grado dos) en dos variables (las coordenadas) igualado a 0. Al tomar cónicas en diferentes posiciones y con distintos parámetros se obtendrán diversos polinomios. La pregunta que guía este capítulo es la pregunta inversa:

¿Será cierto que cualquier polinomio cuadrático en dos variables define una cónica? Si es así, ¿cómo saber qué tipo de cónica es y cuál es su descripción geométrica?

Por su parte, en el Capítulo 3 estudiamos diferentes grupos de transformaciones geométricas. Ellos nos darán la herramienta y el lenguaje para atacar y entender el problema que nos acabamos de plantear.

4.1 ¿Qué es clasificar?

Muy en general, "clasificar" es describir o enumerar las clases de equivalencia de un conjunto de objetos de acuerdo con ciertos "criterios" que definen una relación de equivalencia entre ellos. En nuestro caso, de lo que depende una "clasificación" es de determinar un conjunto de objetos geométricos (triángulos, digamos, como ejemplo para fijar ideas) y de establecer una noción de equivalencia, es decir, definir las condiciones en que estamos dispuestos a decir que dos objetos son "iguales", o, mejor dicho, equivalentes. En esto último, la noción de grupo de transformaciones que hemos desarrollado es básica. Concretemos.

Una *figura plana* es cualquier subconjunto $\mathsf{F} \subset \mathbb{R}^2$. Sea G un grupo de transformaciones de $\mathbb{R}^2$ (nos interesan principalmente los que ya hemos definido, es decir, G puede ser el grupo afín $\mathbf{Af}(2)$; el general lineal $\mathbf{Gl}(2)$; el grupo de semejanzas; el de homotecias; el de isometrías $\mathbf{Iso}(2)$; el de transformaciones ortogonales $\mathbf{O}(2)$, o el de traslaciones). Decimos que dos figuras F_1 y F_2 son G-*equivalentes* (o bien, que son

equivalentes módulo G) y lo denotamos $F_1 \overset{G}{\sim} F_2$, si existe $g \in G$ tal que $g(F_1) = F_2$. La terminología se puede hacer a veces más simple; por ejemplo, dos figuras son *semejantes* (o *isométricas*) si son equivalentes módulo semejanzas (o, respectivamente, **Iso**(2)-equivalentes).

Veamos primero que ésta es efectivamente una relación de equivalencia, demostración que corresponde puntualmente a que G es un grupo de transformaciones:

i) $\overset{G}{\sim}$ es *refexiva* ($F \overset{G}{\sim} F$); pues $\mathrm{id}(F) = F$ y la función identidad, id, está en G.

ii) $\overset{G}{\sim}$ es *simétrica* ($F_1 \overset{G}{\sim} F_2 \Rightarrow F_2 \overset{G}{\sim} F_1$); pues si existe $g \in G$ tal que $g(F_1) = F_2$ entonces $g^{-1}(F_2) = F_1$ y $g^{-1} \in G$.

iii) $\overset{G}{\sim}$ es *transitiva* ($F_1 \overset{G}{\sim} F_2$ y $F_2 \overset{G}{\sim} F_3 \Rightarrow F_1 \overset{G}{\sim} F_3$); pues si $g_1(F_1) = F_2$ y $g_2(F_2) = F_3$ con $g_1, g_2 \in G$, entonces $(g_2 \circ g_1)(F_1) = F_3$ y $(g_2 \circ g_1) \in G$.

Además, estas relaciones se "anidan" de acuerdo con la contención de grupos; es decir, si $H \subset G$ entonces $F_1 \overset{H}{\sim} F_2 \Rightarrow F_1 \overset{G}{\sim} F_2$. Así, si, por decir algo, dos figuras son isométricas, también serán afinmente equivalentes (pero no necesariamente linealmente equivalentes).

4.1.1 Clasificación de triángulos

Para acabar de entender con un ejemplo concreto, describamos la clasificación de triángulos. Primero hay que definirlos: sea $\mathcal{T}$ el conjunto de triángulos planos, es decir, ternas (no ordenadas) no colineales de puntos en $\mathbb{R}^2$. El Teorema 6.4 nos dice que "hay un sólo triángulo afín", es decir que cualquier par de triángulos son afinmente equivalentes.

Esta única clase de equivalencia afín se parte en muchas subclases cuando vemos los triángulos con "ojos de semejanza". Pues, si alguna de las seis transformaciones afines que mandan un triángulo en otro es una semejanza, entonces éstos tienen los mismos ángulos, e inversamente. Así que el teorema "*dos triángulos son semejantes si y sólo si sus ángulos son iguales*" es un "teorema de clasificación" que dice que hay una clase de semejanza por cada terna (no ordenada) de ángulos α, β, γ que cumple $\alpha > 0, \beta > 0, \gamma > 0$ y $\alpha + \beta + \gamma = \pi$.

De nuevo, las clases de semejanza se parten en clases de isometría de acuerdo con un parámetro real y positivo (el tamaño) que podríamos tomar como el área del triángulo o la longitud de un lado dado. Y podríamos de nuevo partir estas clases de isometría en clases módulo translaciones (y aparece un parámetro de "orientación" y uno de "dirección"), o bien en clases módulo ortogonales (y aparece un parametro de "distancia al origen"). Llegando al extremo de la clasificación con el grupo trivial ($\{\mathrm{id}\}$) donde dos son equivalentes si son estrictamente el mismo.

4.2 Clasificación de cónicas

Primero que nada, un polinomio de segundo grado (o *cuadrático*) en las variables x, y es una expresión

$$P(x, y) = ax^2 + bxy + cy^2 + dx + ey + f,$$

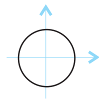
donde los coeficientes a, b, c, d, e, f son números reales, y además al menos uno de a, b, c es distinto de 0 (si no fuera así estaríamos hablando de un polinomio lineal). Los primeros tres términos constituyen su *parte cuadrática,* los siguientes dos su parte *lineal*, y por último f es la parte *constante.* Nótese que al monomio bxy se le considera de grado 2 pues es la suma de los exponentes de las dos variables. A un polinomio cuadrático se le puede pensar también como una función $P : \mathbb{R}^2 \to \mathbb{R}$ donde a un punto dado $\mathbf{x} \in \mathbb{R}^2$ se le asigna el número $P(\mathbf{x})$ que resulta de sustituir en la expresión los valores de sus coordenadas $\mathbf{x}^\top = (x, y)$. Así que podemos hablar de los *ceros del polinomio* P, o la *curva asociada al polinomio* P, como un subconjunto de $\mathbb{R}^2$, a saber:

$$\mathcal{C}(P) = \left\{\mathbf{x} \in \mathbb{R}^2 \mid P(\mathbf{x}) = 0\right\}.$$

Diremos que un subconjunto $\mathcal{C} \subset \mathbb{R}^2$ es una *curva cuadrática* si consiste de los ceros de un polinomio cuadrático, es decir, si para algún polinomio cuadrático P se tiene que $\mathcal{C} = \mathcal{C}(P)$. Por el momento no queremos usar el nombre más común de "cónica", para que quede claro nuestro problema, que consiste en clasificar las curvas cuadráticas de acuerdo con los diferentes grupos geométricos, y en particular, ver si son siempre cónicas. La parte fundamental de esta clasificación consistirá en demostrar que cualquier curva cuadrática es afinmente equivalente a alguna de las siguientes, que por el momento sirven de ejemplos.

4.2.1 Las cónicas canónicas (y algo más)

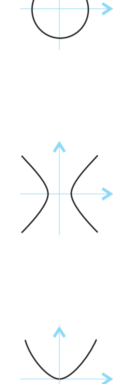
i) El círculo unitario. El polinomio



$$x^2 + y^2 - 1 \tag{i}$$

tiene como ceros el círculo unitario.

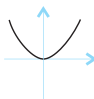
ii) La hipérbola unitaria. El polinomio



$$x^2 - y^2 - 1 \tag{ii}$$

tiene como ceros la hipérbola que llamaremos unitaria.

iii) La parábola canónica. El polinomio



$$x^2 - y \tag{iii}$$

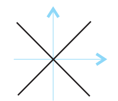
tiene como ceros a una parábola.

Por medio de transformaciones afines estas tres curvas se pueden mandar a muchas otras. Del círculo unitario se obtienen las elipses, y de la hipérbola y la parábola (que hemos escogido por la simpleza de su polinomio) se obtienen todas las demás hipérbolas y parábolas que estudiamos en el Capítulo 2. Pero notemos de una vez que con los polinomios cuadráticos se obtienen nuevos conjuntos:

iv) El círculo imaginario. El polinomio

$$x^2 + y^2 + 1 \tag{iv}$$

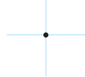
no tiene ningún cero en los reales pues la suma de dos cuadrados es no negativa, más 1 es estrictamente positiva. Es decir, su curva cuadrática es el vacío $\emptyset \subset \mathbb{R}^2$. La llamamos "círculo imaginario" pues fácilmente le encontramos soluciones complejas (por ejemplo $i^2+0^2+1=0$, donde $i^2=-1$) pero se nos sale del contexto de este libro; aunque como polinomio tiene plenos derechos y tendremos entonces que incluirlo en nuestro análisis.



v) Par de rectas. El polinomio

$$x^2 - y^2 \tag{v}$$

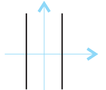
tiene como conjunto de ceros a la unión de las dos rectas $x+y=0$ y $x-y=0$; pues al verlo como producto $(x^2-y^2=(x+y)(x-y))$, si se hace cero entonces alguno de los factores debe ser cero y viceversa.



vi) El círculo de radio cero. Dado por el polinomio

$$x^2 + y^2 \tag{vi}$$

que define al punto 0 por una sola ecuación cuadrática que es el caso limite de círculos cuyo radio se hace cero. También podría llamarse **par de rectas imaginarias** pues se factoriza $(x^2+y^2=(x+iy)(x-iy))$ y se puede ver entonces como dos "rectas complejas" de las que sólo vemos en el plano real al origen.



vii) Rectas paralelas. Por la misma razón que en **(v)**, el polinomio

$$x^2 - 1, \tag{vii}$$

da dos rectas paralelas ($x=1$ y $x=-1$).

viii) Rectas paralelas imaginarias. Dada por

$$x^2 + 1, \tag{viii}$$

que se factoriza $x^2+1=(x+i)(x-i)$.

ix) Recta doble. Dada por



$$x^2. \tag{ix}$$

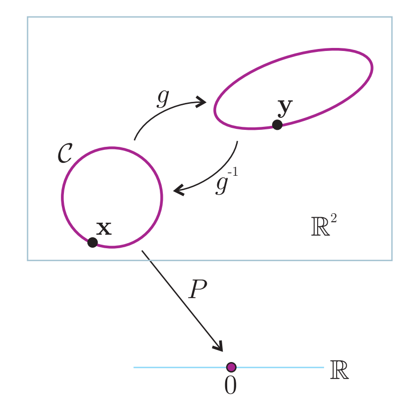
Aunque geométricamente sólo consiste de una recta ($x=0$) se le llama "recta doble" pues el polinomio que la define es diferente.

4.2.2 Equivalencia de polinomios

Para que tenga sentido el problema de clasificar curvas cuadráticas módulo transformaciones afínes, la primera pregunta que debemos responder es si las imágenes afínes de curvas cuadráticas son de nuevo curvas cuadráticas.

Para esto, sean $\mathcal{C}$ una curva cuadrática y $g \in \mathbf{Af}(2)$ una transformación afín. Tenemos entonces que existe un polinomio cuadrático P que define a $\mathcal{C}$, es decir, tal que $\mathcal{C} = \mathcal{C}(P) = \{\mathbf{x} \in \mathbb{R}^2 \mid P(\mathbf{x}) = 0\}$. Afirmamos que

$$g(\mathcal{C}) = \{\mathbf{y} \in \mathbb{R}^2 \mid (P \circ g^{-1})(\mathbf{y}) = 0\}. \tag{4.1}$$



Esta igualdad de conjuntos se demuestra por dos contenciones. Primera ($\subset$): cualquier punto en $g(\mathcal{C})$ es de la forma $g(\mathbf{x})$ con $\mathbf{x} \in \mathcal{C}$; se tiene entonces que $P(\mathbf{x}) = 0$ y por lo tanto

$$(P \circ g^{-1})(g(\mathbf{x})) = P(g^{-1}(g(\mathbf{x}))) = P(\mathbf{x}) = 0,$$

lo cual implica que $g(\mathbf{x})$ está en el conjunto de la derecha.
Y al reves ($\supset$): si tomamos $\mathbf{y}$ tal que $(P \circ g^{-1})(\mathbf{y}) = 0$; sea $\mathbf{x} := g^{-1}(\mathbf{y})$, se tiene entonces que $P(\mathbf{x}) = P(g^{-1}(\mathbf{y})) = 0$ y por tanto que $\mathbf{x} \in \mathcal{C}$, lo cual implica que $\mathbf{y} = g(\mathbf{x}) \in g(\mathcal{C})$. Queda entonces demostrada la igualdad.

Obsérvese que en la demostración anterior sólo se usó que $\mathcal{C}$ estuviera definida como los ceros de una función y que g fuera invertible (una transformación). Ahora veremos que $g(\mathcal{C})$ también es una curva cuadrática. Para esto bastará ver que si P es un polinomio cuadrático y g una transformación afín entonces la función $(P \circ g) : \mathbb{R}^2 \to \mathbb{R}$ también es un polinomio cuadrático (para no estar usando g^{-1}, que sólo complica la notación). Puesto que las dos coordenadas de g son polinomios lineales, y se obtiene una expresión de $P \circ g$ al sustituir estos dos polinomios lineales en las variables de P, es claro que (después de simplificar la expresión) obtendremos un polinomio con monomios de grado a lo más dos. Si los coeficientes de los monomios cuadráticos se cancelaran todos, es decir, si el resultado fuera lineal o constante, tendríamos que al volver a precomponer con g^{-1} se vuelve cuadrático y esto es claramente imposible. Tuvimos que usar entonces que g es una transformación. En resumen, hemos demostrado el siguiente lema.

Lema 4.1 *Sea $\mathcal{C}$ una curva cuadrática y $g \in \mathbf{Af}(2)$, entonces $g(\mathcal{C})$ también es una curva cuadrática. Además, si $\mathcal{C} = \mathcal{C}(P)$ entonces $g(\mathcal{C}) = \mathcal{C}(P \circ g^{-1})$.* □

Hay que observar, antes de definir la equivalencia de polinomios, que al multiplicar un polinomio por una constante $k \neq 0$, sus ceros no cambian. Es decir, que $\mathcal{C}(P) = \mathcal{C}(kP)$ cuando $k \neq 0$; pues entonces $P(\mathbf{x}) = 0 \Leftrightarrow kP(\mathbf{x}) = 0$.

Sea G un subgrupo de $\mathbf{Af}(2)$. Decimos que dos polinomios cuadráticos P_1 y P_2 son G-*equivalentes*, o *equivalentes módulo* G, escrito $P_1 \overset{G}{\sim} P_2$, si existen $g \in G$ y $k \in \mathbb{R}$, con $k \neq 0$, tales que $kP_1 = P_2 \circ g$. Que ésta es una relación de equivalencia se deja como ejercicio.

De lo anterior se obtiene que polinomios cuadráticos G-equivalentes definen curvas G-equivalentes. El inverso, aunque es cierto, aún no lo podemos demostrar, pues incluye que si dos polinomios cuadráticos definen la misma curva (tienen los mismos ceros) entonces difieren por una constante y para esto (en general) hay que incluir los números complejos. Lo que sí podemos hacer es enunciar el teorema fundamental de este capítulo, cuya demostración nos llevará gran parte del mismo.

Teorema 4.2 *Sea* P *un polinomio cuadrático en dos variables* x, y. *Entonces* P *es afinmente equivalente a uno y sólo uno de los polinomios (i), (ii),..., (ix).*

Una vez establecido este teorema, la clasificación afín de curvas cuadráticas será inmediata.

Corolario 4.3 *Hay ocho clases de equivalencia afín de curvas cuadráticas que corresponden a los ocho dibujos de la Sección 4.2.1.*

Nótese que redujimos a ocho pues el vacío aparece en dos tipos de polinomios.

EJERCICIO 4.1 Demuestra que para un subgrupo G de $\mathbf{Af}(2)$, la relación de ser G-*equivalentes* es una relación de equivalencia en los polinomios cuadráticos en dos variables.

4.3 Reducción de polinomios cuadráticos

En esta sección veremos cómo reducir (simplificar) un polinomio cuadrático cualquiera mediante cambios de coordenadas afines. Para empezar, valdrá la pena simplificar la notación de tal polinomio usando el lenguaje matricial y vectorial que hemos desarrollado. A su vez, esto nos dejará ver con nitidez cómo diferentes tipos de transformaciones afines afectan el polinomio.

El polinomio general de segundo grado en dos variables se puede siempre escribir como

$$P(x, y) = ax^2 + 2bxy + cy^2 + dx + ey + f, \tag{4.2}$$

donde la única diferencia con la forma anterior es que hemos puesto un 2 al coeficiente del término mixto (el monomio en xy) por razones que serán evidentes muy pronto. Si consideramos el vector variable $\mathbf{x}^\top = (x, y)$ y definimos una matriz A y un vector $\mathbf{k}$ como

$$A := \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} d \\ e \end{pmatrix},$$

el polinomio P se reescribe como

$$P(\mathbf{x}) = \mathbf{x} \cdot A\mathbf{x} + \mathbf{k} \cdot \mathbf{x} + f, \tag{4.3}$$

que llamaremos la *expresión vectorial* de P, pues

$$\begin{aligned}
\mathbf{x} \cdot A\mathbf{x} &= \begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} ax + by \\ bx + cy \end{pmatrix} \\
&= x(ax + by) + y(bx + cy) = ax^2 + 2bxy + cy^2,
\end{aligned}$$

y

$$\mathbf{k} \cdot \mathbf{x} = \begin{pmatrix} d \\ e \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = dx + ey.$$

Debe quedar claro entonces que anteponer el 2 al coeficiente mixto fue para que la expresión de la matriz A quede más simple. Así la matriz A, que debemos enfatizar que es *simétrica* (es decir, $A = A^\top$), representa a la parte cuadrática, mientras que el vector $\mathbf{k}$ a la parte lineal (la constante f quedó tal cual). Por ejemplo, el polinomio $x^2 + 4xy + 3y^2 + 5x - y + 2$ queda determinado por su parte constante 2, la matriz

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \quad \text{y el vector} \quad \mathbf{k} = \begin{pmatrix} 5 \\ -1 \end{pmatrix}.$$

Nuestra definición de *polinomio cuadrático* será ahora una expresión (4.3) con

$$A = A^\top \neq 0.$$

Nótese que al pedir a la matriz A que sea simétrica obtenemos que la expresión vectorial de P (4.3) es única; se pudo haber tomado cualquier otra matriz con tal de que sus elementos no diagonales sumaran $2b$ pero la más natural, y que resulta única, es la simétrica y esto será importante. Una de las grandes ventajas de usar la expresión vectorial (4.3) en vez de la expresión en variables (4.2) es que la vectorial se escribe igual cuando hay más variables; sólo cambian las dimensiones de los elementos, y da la idea clara de que se pueden generalizar nuestros argumentos. Por ejemplo, para tres variables, tanto el vector variable $\mathbf{x}$ y el constante $\mathbf{k}$ tienen tres entradas, y la matriz A es simétrica de orden 3×3; pero en la expresión en variables faltan cuatro monomios a considerar (véase ejercicio abajo). Sin embargo, esta posibilidad no la usaremos por el momento, y seguiremos con dos variables.

Veremos ahora cómo se afecta la expresión vectorial del polinomio cuadrático P al precomponer con diversas transformaciones afines, empezando por las más simples.

EJERCICIO 4.2 Encuentra la matriz simétrica A y el vector constante $\mathbf{k}$ que dan la expresión vectorial de los siguientes polinomios cuadráticos:

$$x^2 + 2y^2 - 6x + 4y + 3$$
$$9x^2 + 24xy + 16y^2 - 2x - 86y + 39$$
$$2xy - 6x - 4y - 4.$$

EJERCICIO 4.3 Da una expresión general para un polinomio cuadrático en tres variables x, y, z y luego define una expresión vectorial para él.

4.3.1 Traslaciones (cómo encontrar el centro)

Consideremos la traslación $g(\mathbf{x}) = \mathbf{x} + \mathbf{h}$, para cualquier vector $\mathbf{h} \in \mathbb{R}^2$, y veamos cómo se escribe el polinomio cuadrático $P \circ g$, donde P está descrito por (4.3) (es decir, donde $P(\mathbf{x}) = \mathbf{x} \cdot A\mathbf{x} + \mathbf{k} \cdot \mathbf{x} + f$).

$$\begin{aligned}(P \circ g)(\mathbf{x}) &= P(\mathbf{x}+\mathbf{h}) = (\mathbf{x}+\mathbf{h}) \cdot A(\mathbf{x}+\mathbf{h}) + \mathbf{k} \cdot (\mathbf{x}+\mathbf{h}) + f \\ &= (\mathbf{x}+\mathbf{h}) \cdot (A\mathbf{x} + A\mathbf{h}) + \mathbf{k} \cdot \mathbf{x} + \mathbf{k} \cdot \mathbf{h} + f \\ &= \mathbf{x} \cdot A\mathbf{x} + \mathbf{x} \cdot A\mathbf{h} + \mathbf{h} \cdot A\mathbf{x} + \mathbf{h} \cdot A\mathbf{h} + \mathbf{k} \cdot \mathbf{x} + \mathbf{k} \cdot \mathbf{h} + f \\ &= \mathbf{x} \cdot A\mathbf{x} + \mathbf{x} \cdot A\mathbf{h} + \mathbf{h} \cdot A\mathbf{x} + \mathbf{k} \cdot \mathbf{x} + (\mathbf{h} \cdot A\mathbf{h} + \mathbf{k} \cdot \mathbf{h} + f) \\ &= \mathbf{x} \cdot A\mathbf{x} + (\mathbf{x} \cdot A\mathbf{h} + \mathbf{h} \cdot A\mathbf{x} + \mathbf{k} \cdot \mathbf{x}) + P(\mathbf{h}). \end{aligned} \tag{4.4}$$

Nótese que los tres sumandos de enmedio corresponden a la parte lineal, que podemos simplificar aún más usando dos lemas generales.

Lema 4.4 *Dadas dos matrices* A *y* B *que se puedan multiplicar, se tiene* $(AB)^\top = B^\top A^\top$.

Demostración. Ver ejercicios 3.86 y 4.4 (abajo). □

Lema 4.5 *Sea* A *una matriz simétrica (es decir,* $A = A^\top$*) entonces para cualquier par de vectores* $\mathbf{x}, \mathbf{y}$ *se tiene que* $\mathbf{x} \cdot A\mathbf{y} = A\mathbf{x} \cdot \mathbf{y}$.

Demostración. Puesto que $\mathbf{x} \cdot A\mathbf{y} = \mathbf{x}^\top A\mathbf{y}$ y el transpuesto de un número real (matriz 1×1) es él mismo, se tiene por el lema anterior que

$$\begin{aligned}\mathbf{x} \cdot A\mathbf{y} &= \mathbf{x}^\top A\mathbf{y} = (\mathbf{x}^\top A\mathbf{y})^\top = (A\mathbf{y})^\top (\mathbf{x}^\top)^\top \\ &= \mathbf{y}^\top A^\top \mathbf{x} = \mathbf{y}^\top A\mathbf{x} = \mathbf{y} \cdot A\mathbf{x} = A\mathbf{x} \cdot \mathbf{y}.\end{aligned}$$

□

Por este lema, donde es fundamental que A sea simétrica, podemos simplificar la parte lineal de (4.4) para obtener

$$\mathbf{x} \cdot A\mathbf{h} + \mathbf{h} \cdot A\mathbf{x} + \mathbf{k} \cdot \mathbf{x} = 2\,(A\mathbf{h} \cdot \mathbf{x}) + \mathbf{k} \cdot \mathbf{x} = (2A\mathbf{h} + \mathbf{k}) \cdot \mathbf{x},$$

de donde concluimos que

$$(P \circ g)(\mathbf{x}) = P(\mathbf{x} + \mathbf{h}) = \mathbf{x} \cdot A\mathbf{x} + (2A\mathbf{h} + \mathbf{k}) \cdot \mathbf{x} + P(\mathbf{h}).$$

De tal manera que si podemos encontrar $\mathbf{h} \in \mathbb{R}^2$ que cumpla

$$2A\mathbf{h} + \mathbf{k} = 0$$

habremos encontrado una traslación que elimine la parte lineal de P. En caso de que tal $\mathbf{h}$ exista, se le llama el *centro* de la curva cuadrática (o de la cónica) asociada a P. En particular, podemos concluir con el siguiente lema.

Lema 4.6 *Sea* $P(\mathbf{x}) = \mathbf{x} \cdot A\mathbf{x} + \mathbf{k} \cdot \mathbf{x} + f$ *un polinomio cuadrático* ($A = A^\top$) *tal que* $\det(A) \neq 0$*. Entonces si definimos* $\mathbf{c} := -(1/2)A^{-1}\mathbf{k}$*, se tiene que* $\mathbf{c}$ *es el centro de* $\mathcal{C}\,(P)$ *y*

$$P(\mathbf{x} + \mathbf{c}) = \mathbf{x} \cdot A\mathbf{x} + P(\mathbf{c}).$$

□

EJERCICIO 4.4 Demuestra que $(ABC)^\top = C^\top B^\top A^\top$, donde A, B, C son tres matrices que se pueden multiplicar.

EJERCICIO 4.5 Encuentra los centros, si es que tienen, de las curvas asociadas a algunos de los polinomios siguientes:

$$\begin{gathered} x^2 + 2y^2 - 6x + 4y + 3 \\ 9x^2 + 24xy + 16y^2 - 2x - 86y + 39 \\ xy - 3x - 2y - 2 \\ -6x^2 + 24xy + y^2 - 12x - 26y - 161 \\ 9x^2 - 4xy + 6y^2 - 58x + 24y + 59 \\ x^2 + 2xy + y^2 + 2x + 2y - 1. \end{gathered}$$

4.3.2 Rotaciones (cómo encontrar los ejes)

Ya vimos que las traslaciones afectan la parte lineal de los polinomios cuadráticos. Consideremos ahora el caso en que g sea una transformación lineal $g(\mathbf{x}) = B\mathbf{x}$ (donde

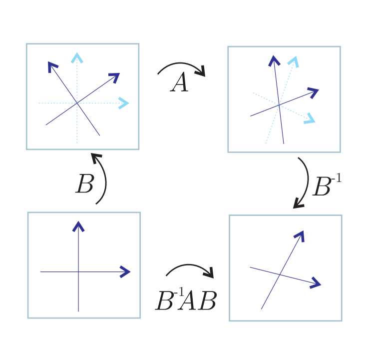
$B \in \mathbf{Gl}(2)$). Si P es el polinomio cuadrático general (4.3) entonces

$$\begin{aligned} (P \circ g)(\mathbf{x}) &= P(B\mathbf{x}) = (B\mathbf{x}) \cdot A(B\mathbf{x}) + \mathbf{k} \cdot (B\mathbf{x}) + f \\ &= (B\mathbf{x})^\top A(B\mathbf{x}) + \mathbf{k}^\top B\mathbf{x} + f \\ &= (\mathbf{x}^\top B^\top)A(B\mathbf{x}) + (\mathbf{k}^\top B)\mathbf{x} + f \\ &= \mathbf{x} \cdot (B^\top AB)\mathbf{x} + (B^\top \mathbf{k}) \cdot \mathbf{x} + f. \end{aligned} \tag{4.5}$$

Obsérvese primero que la nueva matriz $B^\top AB$ que define la parte cuadrática sigue siendo simétrica pues $(B^\top AB)^\top = B^\top A^\top \left(B^\top\right)^\top = B^\top AB$, así que la expresión de $P \circ g$ dada arriba es válida. Simplificar la parte cuadrática de P corresponde entonces a encontrar una matriz B tal que $B^\top AB$ sea más simple; por ejemplo *diagonal,* es decir, con ceros fuera de la diagonal, pues entonces el polinomio $P \circ g$ no tendrá término mixto xy. Veremos, como resultado principal de esta sección, que esto siempre se puede lograr.

Hasta aquí sólo hemos usado que B es una matriz (ni siquiera usamos que fuera no singular, con determinante no nulo), pero si pedimos además que sea ortogonal entonces tenemos que $B^\top = B^{-1}$ y, en este caso, la matriz de la parte cuadrática se transforma en la matriz $B^\top AB = B^{-1}AB$ que tiene un significado geométrico muy claro. Es la matriz que expresa la función A en la base de las columnas de B. Veamos.

Sean $\mathbf{u}$ y $\mathbf{v}$ las columnas de B, es decir $B = (\mathbf{u}, \mathbf{v})$. Si $\mathbf{x}^\top = (x,y)$ entonces $B\mathbf{x} = x\,\mathbf{u} + y\,\mathbf{v}$; y podemos pensar a $B\mathbf{x}$ como el vector (con coordenadas repecto a la base canónica) que tiene coordenadas (x, y) en la base $\mathbf{u},\mathbf{v}$. Por el contrario, dado cualquier $\mathbf{y} \in \mathbb{R}^2$, $B^{-1}\mathbf{y}$ nos da los coeficientes para escribir a $\mathbf{y}$ como combinación de la base $\mathbf{u},\mathbf{v}$, pues es la solución del sistema $B\mathbf{x} = \mathbf{y}$. Así que $B^{-1}AB$ es la matriz que describe a A, pensada como función, pero en términos de la base $\mathbf{u},\mathbf{v}$; por ejemplo, $\left(B^{-1}AB\right)(1,0)^\top = B^{-1}A\left(B\,(1,0)^\top\right) = B^{-1}(A\mathbf{u})$ que da, como habíamos quedado, las coordenadas de $A\mathbf{u}$ en la base $\mathbf{u},\mathbf{v}$ y los datos que le dimos, $(1,0)$, son las coordenadas de $\mathbf{u}$ en la base $\mathbf{u},\mathbf{v}$. Quizá este razonamiento general no convenza aún, pero no importa, pues en nuestro caso lo podemos concretar aún más.



Supongamos, como es el caso en que estamos interesados, que $B \in \mathbf{O}(2)$, es decir que $BB^\top = I$. Y sean $\mathbf{u}$ y $\mathbf{v}$ las columnas de B que forman una base ortonormal. Supongamos además que A actúa de manera muy sencilla en $\mathbf{u}$ y en $\mathbf{v}$: que simplemente los alarga por factores λ y μ respectivamente, es decir que $A\mathbf{u} = \lambda\,\mathbf{u}$ y $A\mathbf{v} = \mu\,\mathbf{v}$. Entonces, expresar a A (como función lineal) pero en términos de la base B debía ser la matriz

$$\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}.$$

Veamos si corresponde a $B^\top AB$:

$$\begin{aligned} B^\top AB &= (u,v)^\top A(u,v) = \begin{pmatrix} u^\top \\ v^\top \end{pmatrix} (Au, Av) = \begin{pmatrix} u^\top Au & u^\top Av \\ v^\top Au & v^\top Av \end{pmatrix} \\ &= \begin{pmatrix} u\cdot Au & u\cdot Av \\ v\cdot Au & v\cdot Av \end{pmatrix} = \begin{pmatrix} u\cdot \lambda u & u\cdot \mu v \\ v\cdot \lambda u & v\cdot \mu v \end{pmatrix} \\ &= \begin{pmatrix} \lambda\,(u\cdot u) & \mu\,(u\cdot v) \\ \lambda\,(v\cdot u) & \mu\,(v\cdot v) \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}. \end{aligned} \tag{4.6}$$

Efectivamente, y con esto nos basta, pues si encontramos una matriz B que satisfaga lo anterior (que "diagonalice" la matriz A) podremos eliminar el término mixto del polinomio P para acercarnos a los polinomios canónicos.

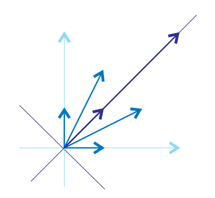
Valores propios y vectores propios

Sea A una matriz cuadrada cualquiera. Se dice que un vector v es *vector propio* de A con *valor propio* $\lambda \in \mathbb{R}$, si $v \neq 0$ y se tiene que $Av = \lambda v$. Nuestro problema será encontrar *parejas propias* (es decir, λ, v valor y vector propio correspondientes). A los vectores propios y valores propios también se les conoce como *eigenvectores* y *eigenvalores* por su nombre en alemán.

Un vector propio de una matriz es un vector donde la matriz (por más complicada que ésta sea), como función, actúa de una manera muy sencilla, simplemente lo alarga. Nótese además que eliminamos el 0 de los posibles vectores propios pues éste sería tal para cualquier "valor propio"; sin embargo el 0 ($\in \mathbb{R}$) sí puede ser valor propio si la matriz anula algún vector no nulo.

Antes de abordar el problema de encontrar vectores propios de manera algebraica, veamos un **ejemplo** sencillo donde geométricamente es fácil encontrarlos. Consideremos la matriz

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$



Puesto que bajo esta función el eje x gira hacia arriba y el eje y hacia la derecha, el cuadrante positivo se apachurra; las líneas en él cercanas al eje x giran en dirección contraria a las cercanas al eje y y entonces debe haber una línea que se queda fija. Por simetría (pasa escencialmente lo mismo en los dos extremos) esta línea debe ser la de enmedio; efectivamente, el $(1,1)$ es vector propio de A con valor propio 3. Los otros dos cuadrantes se abren como abanico y de nuevo su línea "media" se queda fija; el vector $(-1,1)$ es vector propio de A con valor propio 1. Entonces, en la base rotada $45°$ esta transformación consiste simplemente en expander por un factor 3 el primer eje y dejar fijo el ortogonal.

Otro ejemplo geométricamente claro son las reflexiones: tienen vectores propios con valor propio 1 en la línea espejo y vectores propios con valor propio -1 en su ortogonal.

Veamos primero un resultado muy general que agrupa naturalmente a los vectores propios.

Lema 4.7 *Si $\mathbf{u}$ y $\mathbf{v}$ son vectores propios de A con el mismo valor propio $\lambda \in \mathbb{R}$, entonces cualquier combinación lineal no trivial de ellos también es vector propio.*

Demostración. Si $A\mathbf{u} = \lambda\,\mathbf{u}$ y $A\mathbf{v} = \lambda\,\mathbf{v}$, entonces para cualquier par de coeficientes $\alpha, \beta \in \mathbb{R}$ se tiene que

$$A(\alpha\,\mathbf{u} + \beta\,\mathbf{v}) = \alpha(A\mathbf{u}) + \beta(A\mathbf{v}) = \alpha(\lambda\,\mathbf{u}) + \beta(\lambda\,\mathbf{v}) = \lambda(\alpha\,\mathbf{u} + \beta\,\mathbf{v})$$

y entonces, si $\alpha\,\mathbf{u} + \beta\,\mathbf{v} \neq 0$, es vector propio de A con valor propio λ. □

Así que los vectores propios vienen en "paquetes" que son subespacios (líneas por el origen es lo más común) asociados a los valores propios. Los conjuntos de vectores propios asociados a un valor propio se llaman *espacios propios* o *eigenespacios*; en ellos, la matriz actúa como una homotecia. Por ejemplo, la matriz

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

tiene como espacio propio del valor propio 2 al plano xy y como espacio propio del valor propio 1 al eje z.

Supongamos que $\lambda, \mathbf{v}$ es una pareja propia de la matriz A (en particular $\mathbf{v} \neq 0$), entonces

$$\begin{aligned} & A\mathbf{v} = \lambda\,\mathbf{v} \\ \Leftrightarrow\quad & A\mathbf{v} - \lambda\,\mathbf{v} = 0 \\ \Leftrightarrow\quad & (A - \lambda\,I)\,\mathbf{v} = 0. \end{aligned}$$

Esta última ecuación nos dice que la transformación $(A - \lambda\,I)$ no es biyectiva pues manda a un vector no cero en el cero. Y como $\mathbf{v} \neq 0$, esto último sucede (regresando formalmente al caso de dimensión 2 que sí hemos demostrado, aunque será cierto en general) si y sólo si

$$\det(A - \lambda\,I) = 0.$$

Esto demuestra el siguiente lema.

Lema 4.8 *Para cualquier matriz A (de 2×2):*

$$\lambda \textit{ es un valor propio de } A \Leftrightarrow \det(A - \lambda I) = 0.$$

□

Lema 4.9 *Si A es simétrica 2×2 entonces tiene dos valores propios λ_1 y λ_2 en $\mathbb{R}$.*

Demostración. Puesto que A es simétrica, $A = A^\top$, la podemos escribir

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Consideremos ahora a λ como una variable o incógnita y entonces tenemos

$$\begin{aligned} \det(A - \lambda I) &= \det\left(\begin{pmatrix} a & b \\ b & c \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) \\ &= \det\begin{pmatrix} a-\lambda & b \\ b & c-\lambda \end{pmatrix} = (a-\lambda)(c-\lambda) - b^2 \\ &= \lambda^2 - (a+c)\lambda + (ac - b^2). \end{aligned}$$

Que resulta ser un polinomio de segundo grado en la variable λ.

Este polinomio es tan importante que tiene nombre: $\det(A - \lambda I)$ es el *polinomio característico* de la matriz A.

Nos interesa, según el lema anterior, determinar para cuáles λ se anula este polinomio, es decir, cuáles son sus raíces. Para esto apliquemos la fórmula clásica "del chicharronero" (página 77), la cual expresa las raíces en términos de los coeficientes y en nuestro caso da:

$$\begin{aligned} \frac{(a+c) \pm \sqrt{(a+c)^2 - 4(ac-b^2)}}{2} &= \frac{(a+c) \pm \sqrt{a^2 + 2ac + c^2 - 4ac + 4b^2}}{2} \\ = \frac{(a+c) \pm \sqrt{a^2 - 2ac + c^2 + 4b^2}}{2} &= \frac{(a+c) \pm \sqrt{(a-c)^2 + 4b^2}}{2}. \end{aligned}$$

Puesto que $(a-c)^2 + 4b^2 \geq 0$ por ser la suma de dos cuadrados, entonces la raíz cuadrada es real y $\det(A - \lambda I) = 0$ tiene dos soluciones reales, que vienen de tomar $+$ o $-$ en la expresión anterior. Explícitamente, sean

$$\lambda_1 = \frac{(a+c) + \sqrt{(a-c)^2 + 4b^2}}{2}, \quad \lambda_2 = \frac{(a+c) - \sqrt{(a-c)^2 + 4b^2}}{2}$$

las dos raíces; de tal manera que podemos concluir, por el Lema 4.8, que λ_1 y λ_2 son valores propios de A y que su polinomio característico se escribe

$$\det(A - \lambda I) = (\lambda - \lambda_1)(\lambda - \lambda_2).$$

□

Obsérvese que en la demostración fue crucial que $A^\top = A$; pues en general este lema no es cierto (véanse los ejercicios): no toda matriz A de 2×2 tiene valores propios, como no todo polinomio cuadrático en una variable tiene raíces reales.

Además, de la demostración podemos concluir (para matrices simétricas) que los dos valores propios coinciden si y solamente si $a = c$ y $b = 0$ (pues se debe tener que $(a-c)^2 + 4b^2 = 0$); es decir, los dos valores propios coinciden si y sólo si la matriz original es la de una homotecia. Tenemos entonces:

Corolario 4.10 *Sea* A *simétrica* 2×2 *y sean* λ_1, λ_2 *sus dos valores propios. Entonces, sus valores propios coinciden* ($\lambda_1 = \lambda_2$) *si y sólo si* $A = \lambda_1 I$. *Y en este caso, cualquier vector* $\mathbf{v} \neq \mathbf{0}$ *es vector propio.* □

Corolario 4.11 *Sea* A *simétrica* 2×2. *Entonces existe una base* $\mathbf{u}, \mathbf{v}$ *de* $\mathbb{R}^2$ *donde* $\mathbf{u}$ *y* $\mathbf{v}$ *son vectores propios de* A.

Demostración. Por el Lema 4.9 sabemos que A tiene dos valores propios, λ_1 y λ_2. Si éstos fueran iguales, por el corolario anterior cualquier base funciona. Si fueran diferentes ($\lambda_1 \neq \lambda_2$), entonces por la definición de valor propio existen vectores propios $\mathbf{u}, \mathbf{v}$ distintos de $\mathbf{0}$ correspondientes a los valores propios λ_1 y λ_2 respectivamente. Estos dos vectores no pueden ser paralelos pues por el Lema 4.8 sus valores propios coincidirían. Entonces forman una base de $\mathbb{R}^2$. □

Un ejemplo Vale la pena hacer un ejemplo concreto para fijar ideas. Consideremos la matriz simétrica

$$A = \begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix}.$$

Para encontrar sus valores propios hay que resolver su polinomio característico que es

$$\begin{aligned} \det(A - \lambda I) &= \det\left(\begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right) \\ &= \det\begin{pmatrix} 2-\lambda & 2 \\ 2 & -1-\lambda \end{pmatrix} \\ &= (2-\lambda)(-1-\lambda) - 4 \\ &= \lambda^2 - \lambda - 6. \end{aligned}$$

La fórmula nos dice que sus raíces son

$$\lambda_1 = (1 + \sqrt{1+24})/2 = 3 \quad \text{y} \quad \lambda_2 = (1-5)/2 = -2,$$

lo cual implica que

$$\det(A - \lambda I) = (\lambda - 3)(\lambda + 2).$$

Aunque en este caso pudo haber sido resuelto sin usar la fórmula, ya tenemos los dos valores propios. Para encontrar los vectores propios correspondientes, habrá que encontrar una solución no trivial de los sistemas $(A - \lambda_i I)\mathbf{x} = \mathbf{0}$. Para $\lambda_1 = 3$ tenemos

$$(A - \lambda_1 I)\mathbf{x} = \begin{pmatrix} -1 & 2 \\ 2 & -4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x + 2y \\ 2x - 4y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

y una de sus (múltiples) soluciones no triviales es $\mathbf{u}^\top := (2, 1)$. Compruébese directamente que $A\mathbf{u} = 3\mathbf{u}$ como reza la teoría. Para el otro eigenvalor, $\lambda_2 = -2$, el sistema es

$$(A - \lambda_2 I)\mathbf{x} = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4x + 2y \\ 2x + y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

y una solución no trivial es $\mathbf{v}^\top = (-1, 2)$.

No sólo obtuvimos que los vectores propios $\mathbf{u}$ y $\mathbf{v}$ forman una base de $\mathbb{R}^2$, sino mucho más: que son ortogonales. Pero esto, como veremos, no es coincidencia.

EJERCICIO 4.6 Demuestra que la matriz $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ no tiene vectores propios. Da un argumento algebraico y uno geométrico.

EJERCICIO 4.7 Demuestra que la matriz $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ no tiene vectores propios para $b \neq 0$. ¿Puedes dar un argumento geométrico?

Conclusión (diagonalización ortogonal de matrices simétricas)

En nuestro estudio de los vectores propios de las matrices simétricas, saquémosle jugo ahora a lo que ya sabíamos sobre su buen comportamiento respecto al producto interior.

Lema 4.12 *Sea A una matriz simétrica. Si $\lambda_1, \mathbf{u}$ y $\lambda_2, \mathbf{v}$ son pares propios tales que $\lambda_1 \neq \lambda_2$, entonces $\mathbf{u}$ y $\mathbf{v}$ son ortogonales.*

Demostración. Usando el Lema 4.5, tenemos

$$\lambda_1(\mathbf{u} \cdot \mathbf{v}) = (\lambda_1 \mathbf{u}) \cdot \mathbf{v} = A\,\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot A\,\mathbf{v} = \mathbf{u} \cdot (\lambda_2 \mathbf{v}) = \lambda_2(\mathbf{u} \cdot \mathbf{v}),$$

de donde

$$(\lambda_1 - \lambda_2)(\mathbf{u} \cdot \mathbf{v}) = 0.$$

Si $\lambda_1 \neq \lambda_2$, esto implica que $\mathbf{u} \cdot \mathbf{v} = 0$. □

Podemos entonces concluir.

Teorema 4.13 *Sea A una matriz simétrica 2×2. Existe una rotación $B \in \boldsymbol{O}(2)$ tal que $B^\top A B$ es diagonal, es decir, tal que*

$$B^\top A B = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

donde λ_1 y λ_2 son los valores propios de A.

Demostración. Por el Lema 4.9, A tiene valores propios $\lambda_1, \lambda_2 \in \mathbb{R}$. Si $\lambda_1 = \lambda_2$ entonces A ya es diagonal por el Corolario 4.10, y B puede tomarse entonces como la identidad. Si $\lambda_1 \neq \lambda_2$, sean $\mathbf{u}$ y $\mathbf{v}$ vectores propios respectivos. Por el Lema 4.12, $\mathbf{u}$ y $\mathbf{v}$ son ortogonales, entonces $\mathbf{v}$ es paralelo a $\mathbf{u}^\perp$; y por el Lema 4.7, $\mathbf{u}^\perp$ también es vector propio correspondiente a λ_2. Sea

$$B = \frac{1}{|\mathbf{u}|} \left(\mathbf{u}, \mathbf{u}^\perp\right);$$

puesto que $\mathbf{u} \neq 0$, claramente B es la matriz de una rotación. Y con un cálculo análogo al que hicimos en (4.6) se concluye el teorema. □

La teoría que hemos desarrollado se resume en una receta muy simple.

Receta para diagonalizar matrices simétricas. El **ingrediente** es una matriz simétrica $A = A^\top$ de 2×2.

El **primer paso** consiste en resolver su polinomio característico $\det(A - \lambda I)$.

Una vez que se conocen $\lambda_1, \lambda_2 \in \mathbb{R}$ tales que $\det(A - \lambda I) = (\lambda - \lambda_1)(\lambda - \lambda_2)$, **procédase** a encontrar $\mathbf{u} \neq 0$ tal que $(A - \lambda_1 I)\mathbf{u} = 0$ (esto le resultará muy fácil si recuerda las propiedades del compadre ortogonal, pues basta aplicarlo a un renglón de $(A - \lambda_1 I)$).

Y todo ya está cocinado para **declarar** $B = (1/|\mathbf{u}|)(\mathbf{u}, \mathbf{u}^\perp)$. El resultado será que $B^\top AB$ es la matriz diagonal con entradas λ_1 y λ_2.

EJERCICIO 4.8 Concluye la demostración del Teorema 4.13.

EJERCICIO 4.9 Encuentra la matriz B de una rotación que diagonalize las siguientes matrices simétricas:

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} -6 & 12 \\ 12 & 1 \end{pmatrix}$$
$$\begin{pmatrix} 9 & -2 \\ -2 & 6 \end{pmatrix} \quad \begin{pmatrix} -7 & -6 \\ -6 & 2 \end{pmatrix} \quad \begin{pmatrix} 9 & 12 \\ 12 & 16 \end{pmatrix}.$$

Calcula $B^\top AB$, donde A es una de las anteriores y B la rotación correspondiente.

EJERCICIO 4.10 Demuestra que si una matriz A (no necesariamente simétrica) tiene dos valores propios distintos entonces existe una matriz $B \in \mathbf{Gl}(2)$ tal que $B^{-1}AB$ es diagonal.

4.4 Clasificación de curvas cuadráticas

Hemos estudiado cómo afectan las traslaciones y las transformaciones ortogonales a los polinomios cuadráticos; ahora juntaremos nuestros resultados para dar su clasificación

isométrica. Pero empecemos con un ejemplo concreto para fijar las ideas, técnicas y resultados que obtuvimos en la sección anterior.

4.4.1 Ejemplo

Consideremos el polinomio cuadrático

$$P(x, y) = 8x^2 + 12xy + 17y^2 - 16x - 12y - 12;$$

queremos saber qué curva define y dónde está. Para pasarlo a su forma vectorial, $P(\mathbf{x}) = \mathbf{x} \cdot A\mathbf{x} + \mathbf{k} \cdot \mathbf{x} + f$, definimos $f = -12$,

$$A = \begin{pmatrix} 8 & 6 \\ 6 & 17 \end{pmatrix} \qquad \text{y} \qquad \mathbf{k} = \begin{pmatrix} -16 \\ -12 \end{pmatrix}.$$

El determinante de A es $\det(A) = 8 \times 17 - 6 \times 6 = 100$ distinto de cero, y por tanto $\mathcal{C}(P)$ tiene centro

$$\mathbf{c} = -A^{-1}\left(\frac{1}{2}\mathbf{k}\right) = \frac{1}{100}\begin{pmatrix} 17 & -6 \\ -6 & 8 \end{pmatrix}\begin{pmatrix} 8 \\ 6 \end{pmatrix} = \frac{1}{100}\begin{pmatrix} 8 \times 17 - 6 \times 6 \\ -8 \times 6 + 6 \times 8 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

De tal manera que si trasladamos la curva $\mathcal{C}(P)$ por el vector $(-1, 0)$ ésta queda definida por el polinomio

$$P_1(\mathbf{x}) = P(\mathbf{x} + \mathbf{c}) = \mathbf{x} \cdot A\mathbf{x} + P(\mathbf{c}) = \mathbf{x} \cdot A\mathbf{x} - 20.$$

Ahora queremos encontrar una rotación B que elimine el término mixto de P_1, y para esto, seguimos la receta de diagonalización. El polinomio caracteristico de A es

$$\det(A - \lambda I) = \det\begin{pmatrix} 8 - \lambda & 6 \\ 6 & 17 - \lambda \end{pmatrix} = \lambda^2 - 25\lambda + 100 = (\lambda - 5)(\lambda - 20)$$

y entonces sus valores propios son $\lambda_1 = 5$ y $\lambda_2 = 20$. Para el primero de ellos, se tiene que

$$A - 5I = \begin{pmatrix} 3 & 6 \\ 6 & 12 \end{pmatrix}$$

y podemos tomar como solución del sistema homogéneo que define a $\mathbf{u}^\top = (2, -1)$, de tal manera que una rotación que diagonaliza a A es

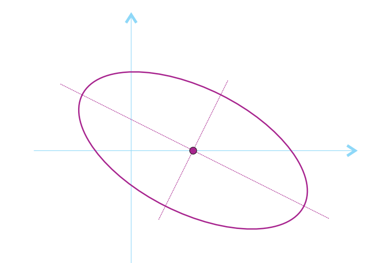
$$B = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}; \quad B^\top AB = \begin{pmatrix} 5 & 0 \\ 0 & 20 \end{pmatrix}.$$

Podemos concluir, por (4.5), que si rotamos el eje x a la línea con dirección $(2,-1)$ el polinomio P_1 toma la forma

$$P_2(\mathbf{x}) = P_1(B\mathbf{x}) = 5x^2 + 20y^2 - 20$$

que, finalmente, es equivalente a

$$P_0(\mathbf{x}) = \frac{1}{5}P_2(\mathbf{x}) = x^2 + 4y^2 - 4.$$



La curva cuadrática de P_0 es claramente una elipse dada por la ecuación canónica

$$\frac{x^2}{4} + y^2 = 1$$

Entonces podemos concluir que $\mathcal{C}(P)$, la que originalmente queríamos describir, es una elipse con centro en $(1,0)$, eje mayor de tamaño 2 en la dirección del vector $(2,-1)$ y eje menor 1.

Ya que tenemos los numeritos calculados y frescos, prosigamos con el ejemplo hasta su final afín. Si al círculo unitario lo expandemos en el eje x un factor 2, y luego componemos con la isometría $B\mathbf{x}+\mathbf{c}$, debemos llegar justo a $\mathcal{C}(P)$. La transformación afín que acabamos de describir tiene la fórmula

$$g\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 2x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{4}{\sqrt{5}}x + \frac{1}{\sqrt{5}}y + 1 \\ -\frac{2}{\sqrt{5}}x + \frac{2}{\sqrt{5}}y \end{pmatrix}.$$

Puesto que dividimos entre 5 y luego entre 4 para obtener la ecuación canónica de la elipse, debe ser cierto, según la teoría que hemos desarrollado, que

$$(P \circ g)(x,y) = 20\left(x^2 + y^2 - 1\right).$$

EJERCICIO 4.11 Ensuciate las manos y haz con cuidado todo el trabajo ("talacha" le llamamos en México) que compruebe nuestra última aseveración; simplemente hay que sustituir fórmulas, expandir y simplificar. O bien, prográmalo y revísalo en algún paquete matemático adecuado.

4.4.2 Clasificación isométrica

Consideremos de nuevo el polinomio cuadrático general

$$P(\mathbf{x}) = \mathbf{x}\cdot A\mathbf{x} + 2\,\mathbf{b}\cdot\mathbf{x} + c,$$

con $A = A^\top \neq 0$; donde el vector constante $\mathbf{b}$ que determina la parte lineal es ahora la mitad del que habíamos tomado antes ($\mathbf{k}$) para evitar el $\frac{1}{2}$ que nos aparecía; nótese también que hemos cambiado el nombre a la constante. Y sea

$$g(\mathbf{x}) = B\mathbf{x} + \mathbf{h}$$

una isometría, es decir, $B \in \mathbf{O}(2)$. Al hacer los cálculos como en las secciones precedentes, se obtiene que

$$\begin{aligned}(P \circ g)(\mathbf{x}) &= \mathbf{x} \cdot \left(B^\top AB\right)\mathbf{x} + 2\,(A\mathbf{h} + \mathbf{b}) \cdot (B\mathbf{x}) + P(\mathbf{h}) \\ &= \mathbf{x} \cdot \left(B^\top AB\right)\mathbf{x} + 2\,B^\top (A\mathbf{h} + \mathbf{b}) \cdot \mathbf{x} + P(\mathbf{h}). \end{aligned} \tag{4.7}$$

Debemos partir nuestro análisis en dos grandes casos de acuerdo con el determinante de la matriz A.

Caso 1: $\det(A) \neq 0$.

Si escogemos a $\mathbf{h}$ como el centro ($\mathbf{h} = -A^{-1}\mathbf{b}$) y a B como una rotación que diagonalice A (bien definida salvo el orden y orientación que demos a los vectores propios), entonces P es isométricamente equivalente a un polinomio de la forma

$$P_1(x, y) = \alpha x^2 + \beta y^2 + \gamma, \tag{4.8}$$

donde α y β son los valores propios de A y γ es el valor del polinomio original evaluado en el centro (nótese que con estos datos bastan).

Ahora bien, tenemos también que α y β son distintos de cero por el siguiente lema que habíamos postergado.

Lema 4.14 *Si A es una matriz simétrica con valores propios α y β, entonces*

$$\det(A) = \alpha\beta$$

Demostración. Por el Teorema de diagonalización, 4.13, sea B una rotación que diagonaliza a A. Entonces como el determinante es multiplicativo se tiene que

$$\begin{aligned}\alpha\beta &= \det\begin{pmatrix}\alpha & 0 \\ 0 & \beta\end{pmatrix} = \det\left(B^\top AB\right) \\ &= \det\left(B^\top\right)\det(A)\det(B) = 1 \times \det(A) \times 1 \\ &= \det(A).\end{aligned}$$

□

Entonces tenemos naturalmente dos subcasos: cuando $\det(A) > 0$ y los signos de α y β coinciden, o bien cuando son distintos y $\det(A) < 0$.

$\underline{\det(A) > 0}$: Tenemos tres posibilidades. Primera, si la costante γ es cero (el centro es parte de la "curva" $\mathcal{C}(P)$) entonces la única solución (el único cero de (4.8)) es $(x, y) = (0, 0)$, pues de otra forma es estrictamente del signo de los eigenvalores. Segunda, si la constante γ es del mismo signo que α y β, entonces la curva es vacia pues no hay soluciones (reales, insistamos, pues se le puede llamar "elipse imaginaria"). Y por último, si el signo de γ es el opuesto a los de α y β, los ceros de P_1 (véase (4.8)) claramente coinciden con las soluciones de la ecuación canónica de la elipse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

donde $a = \sqrt{-\gamma/\alpha}$ y $b = \sqrt{-\gamma/\beta}$.

$\underline{\det(A) < 0}$: Sólo tenemos dos posibilidades. Si $\gamma = 0$, P_1 es una diferencia de cuadrados que se puede factorizar en la forma

$$(a\,x + b\,y)\,(a\,x - b\,y)$$

donde hemos supuesto que $\alpha > 0$ (y entonces $a = \sqrt{\alpha}$ y $b = \sqrt{-\beta}$), cuya curva consiste en dos rectas cuya intersección es el centro. O bien, si $\gamma \neq 0$ podemos escoger el primer vector propio (correspondiente a la variable x) de tal manera que su valor propio α tenga signo contrario a γ, y entonces los ceros de P_1 corresponden a las soluciones de la ecuación canónica de la hipérbola:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

donde $a = \sqrt{-\gamma/\alpha}$ y $b = \sqrt{\gamma/\beta}$.

Caso 2: $\det(A) = 0$.

Puesto que no tenemos la seguridad de eliminar la parte lineal, nos conviene simplificar primero la parte cuadrática (cuidado en los ejercicios, esto modifica la parte lineal). En este caso, por el Lema 4.14, uno de los valores propios es cero, y el otro es distinto de cero pues, si no, por el Corolario 4.10, la matriz A sería la matriz nula que hemos desechado desde el principio. Entonces, podemos diagonalizar (rotar) para que el valor propio correspondiente a x sea el no nulo, y dividiendo entre él obtener que P es isométricamente equivalente a un polinomio de la forma

$$x^2 + 2\,\alpha\,x + \beta\,y + \gamma.$$

Este polinomio siempre se puede reescribir como

$$(x + \alpha)^2 + \beta\,y + \left(\gamma - \alpha^2\right),$$

de tal manera que haciendo el cambio de variable isométrico $x' = x + \alpha$, y olvidando la prima, se simplifica a

$$P_2(x, y) = x^2 + a\,y + b. \qquad (4.9)$$

donde a y b tienen las definiciones obvias. ¿A precomponer con qué traslación corresponde este cambio de variable? Y de nuevo tenemos dos subcasos.

$\underline{a = 0}$: Si $b < 0$ entonces el polinomio (4.9) define (y por tanto $\mathcal{C}(P)$ consiste de) dos rectas paralelas. Si $b = 0$, es una recta doble; y si $b > 0$ consiste de "dos rectas imaginarias" (en $\mathbb{R}^2$ es el conjunto vacío).

$\underline{a \neq 0}$: Con una traslación en el eje y, o equivalentemente, el cambio de variable $y' = y + b/a$ y olvidando la prima, obtenemos que P es isométricamente equivalente al polinomio

$$x^2 + a\,y,$$

que define una parábola.

En resumen, hemos demostrado que cualquier polinomio cuadrático es isométricamente equivalente a alguna de nueve posibles familias canónicas. Enfatizamos lo de familias, pues nótese que (con isometrías) no podemos deshacernos de ciertos parámetros. Por ejemplo, y para describir los casos importantes, en las elipses nos quedaron los parámetros a y b en la ecuación canónica, que podemos suponer positivos (pues como se usan sus cuadrados el signo da lo mismo) y además podemos suponer que $a \geq b$ (pues al rotar $90°$ se intercambian las variables x y y en la ecuación). De tal manera que, como intuitivamente ya sabíamos, hay una clase isométrica de elipses para cada pareja $a \geq b > 0$ (los semiejes major y menor); los casos extremos $a = b > 0$ corresponden a los círculos. En las hipérbolas, no podemos intercambiar las variables (cuyo papel se define respecto a la constante) y tenemos una clase para cada par a, b estrictamente positivo. Por último, las parábolas dependen del parametro $a > 0$ (pues cambiar y por $-y$ es una isometría).

EJERCICIO 4.12 Describe geométricamente algunas de las curvas cuadráticas definidas por los polinomios siguientes (da el centro, la dirección de los ejes y los parámetros o la ecuación canónica correspondiente). Escoge al menos una elipse, una hipérbola y una parábola.

$$\begin{gathered}
-6x^2 + 24xy + y^2 - 12x - 26y - 161 \\
9x^2 - 4xy + 6y^2 - 58x + 24y + 59 \\
x^2 + 2xy + y^2 + 2x + 2y - 1 \\
47x^2 + 32xy - 13y^2 - 252x - 12y + 247 \\
5x^2 + 24xy - 5y^2 + 14x - 34y - 37 \\
66x^2 - 24xy + 59y^2 - 108x - 94y + 1 \\
13x^2 + 10xy + 13y^2 + 42x - 6y - 27 \\
9x^2 + 6xy + 17y^2 + 12x - 28y - 52 \\
-7x^2 - 12xy + 2y^2 + 40x + 20y - 55 \\
9x^2 - 24xy + 16y^2 + 130x - 90y + 175 \\
18x^2 + 48xy + 32y^2 - 29x + 3y - 22 \\
32x^2 + 48xy + 18y^2 + 31x - 8y - 88 \\
-7x^2 + 48xy + 7y^2 + 158x - 6y - 88 \\
-24x^2 - 14xy + 24y^2 + 152x - 164y + 151 \\
7x^2 + 48xy - 7y^2 + 116x - 138y - 348.
\end{gathered}$$

EJERCICIO 4.13 Encuentra un polinomio que defina las siguientes curvas cuadráticas.
a) La elipse con semieje mayor 3 en la dirección $(3, 4)$, semieje menor 2 y centro en $(-1, 2)$.
b) La hiperbóla con semieje principal 4 en la dirección $(2, 1)$, semieje secundario 1 y centro en $(2, 3)$.
c) La parábola con vértice en $(1, 3)$ directriz en la dirección $(1, 4)$ y distancia focal 1.

4.4.3 Clasificación afín y por semejanzas

Para concluir con la clasificación afín que habíamos anunciado al principio de la Sección 4.2, bastará convertir los parámetros que nos quedan (α, β y γ, o bien a y b, en los párrafos anteriores) en 1 o -1. Y es claro que esto debe lograrse alargando (o encogiendo) adecuadamente en los ejes, y saliéndonos entonces del ámbito de las isometrías. Por ejemplo, en el polinomio de la parábola, $P(x, y) = x^2 + ay$ con $a \neq 0$, se obtiene que

$$P\left(x, -\frac{y}{a}\right) = x^2 - y,$$

que equivale a precomponer con la expansión de $-a^{-1}$ en el eje y, y que nos da justo el polinomio de la parábola canónica, (iii) en la Sección 2.1. Éste es el único polinomio con parte lineal de la clasificación isométrica, así que podemos olvidarnos de estos términos.

Si en P_1 (4.8) o en P_2 con $a = 0$, hay un término constante no nulo podemos dividir entre él para hacerlo 1. Y nos queda entonces preocuparnos por los términos cuadráticos puros como αx^2. Obsérvese que, si en vez de x, sustituimos la expresión $a x$ (una nueva a, cualquier $a \neq 0$), se obtiene $\alpha\, (a\, x)^2 = \left(\alpha\, a^2\right) x^2$ de tal manera que el coeficiente puede cambiar, pero nunca cambiará de signo. En términos de matrices, si precomponemos a P_2 (4.9) con una transformación lineal que expanda el eje x por un factor a y el eje y por b, la matriz de la parte cuadrática resultante será:

$$\text{“}B^\top AB\text{"} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = \begin{pmatrix} a^2\alpha & 0 \\ 0 & b^2\beta \end{pmatrix}.$$

Así que lo que debemos hacer es tomar $a = \left(\sqrt{|\alpha|}\right)^{-1}$ y $b = \left(\sqrt{|\beta|}\right)^{-1}$ para obtener una matriz del tipo

$$\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}.$$

Con estas observaciones, es fácil ver que los subcasos de la clasificación isométrica dada en la sección anterior corresponden a los incisos y polinomios de la Sección 2.1, con lo que se completa el teorema de clasificación afín.

Si consideramos por último el grupo de semejanzas, entre las isometrías y las afines, en la matriz “B” última que usamos se debe tener que $a = b$. Entonces sólo podemos lograr que uno de los coeficientes se haga ± 1, y por tanto las elipses (imaginarias o no) y las hipérbolas módulo semejanzas tienen un parámetro que puede ser la razón entre los semiejes o, como habíamos visto en el Capítulo 2, la excentricidad.

EJERCICIO 4.14 ¿Cuál es la matriz de una homotecia que lleve a la parábola dada por $x^2 + ay$, con $a \neq 0$, en la canónica (dada por $x^2 - y$)? Concluye que hay sólo una clase de parábolas módulo semejanzas.

4.5 *Lo que no demostramos

Aunque sí demostramos (detalles más, detalles menos) que cualquier polinomio cuadrático es afinmente equivalente a alguno de los nueve canónicos, no demostramos que estos no son equivalentes entre sí. Es decir, que son justo nueve las clases de equivalencia. Un argumento de tipo topológico va como sigue.

Si un polinomio de una elipse fuera equivalente a uno de una hipérbola, entonces tenemos una transformación afín, *afinidad* debíamos llamarla para resumir,

que manda a la elipse en la hipérbola. Pero esto no es posible porque la hipérbola parte el plano en tres pedazos (que se llaman las "componentes conexas" de su complemento) y la elipse sólo en dos. Para diferenciar a la elipse de la parábola necesitamos usar otra cualidad; la de acotado funciona, pues la elipse lo es mientras que la parábola no. En el fondo estamos usando que las *afinidades* son continuas y que bajo transformaciones continuas estas propiedades se preservan. Para diferenciar a las degeneradas no vacías entre sí, este tipo de argumentos funcionan; y para diferenciarlas con las no degeneradas habrá que usar que estas últimas no contienen rectas y son más que un punto. Sin embargo, con estos argumentos topológicos no podemos diferenciar a los dos polinomios de las curvas vacías, aunque a estos sí de los demás.

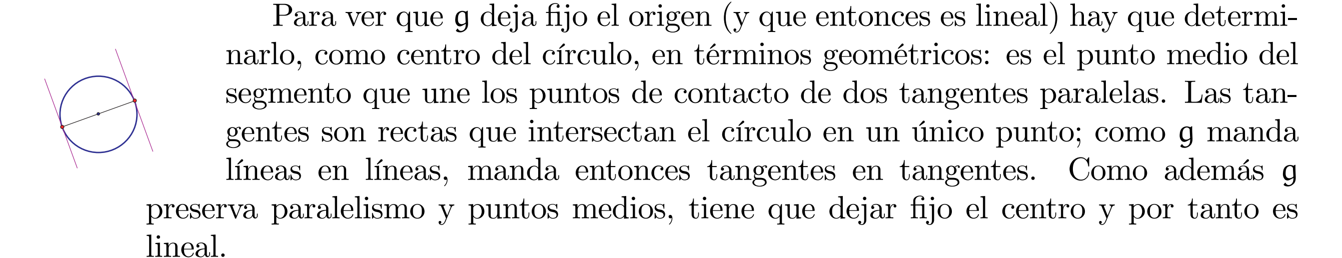
Otra cosa que no demostramos es que los polinomios que definen a una misma cónica son únicos salvo constantes diferentes de cero. Como veremos, esto tiene que ver con su grupo de simetrías afines. Consideremos el caso más simple; el del círculo unitario $\mathbb{S}^1$ cuya matriz asociada es la identidad, pues está definido por el polinomio

$$P(\mathbf{x}) = \mathbf{x} \cdot \mathbf{x} - 1 = \mathbf{x} \cdot I\mathbf{x} - 1.$$

Supongamos que otro polinomio cuadrático P_1 también define el círculo unitario, es decir, que $\mathbb{S}^1 = \mathcal{C}(P_1)$. Queremos demostrar que entonces existe $k \neq 0$ tal que $P_1 = kP$. Por el Teorema de Clasificación Afín y el párrafo precedente sabemos que P_1 y P tienen que ser afinmente equivalentes. Entonces existen $g \in \mathbf{Af}(2)$ y $k \neq 0$ tales que

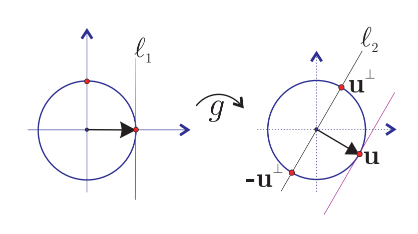
$$P_1 = k(P \circ g).$$

Esto implica que g manda biyectivamente a $\mathbb{S}^1$ en $\mathbb{S}^1$ (arguméntalo). Es decir, que g es una simetría afín del círculo unitario. Ejemplos de este tipo de las g son las transformaciones ortogonales, y ahora demostraremos con argumentos geométricos que son las únicas.



Para ver que g deja fijo el origen (y que entonces es lineal) hay que determinarlo, como centro del círculo, en términos geométricos: es el punto medio del segmento que une los puntos de contacto de dos tangentes paralelas. Las tangentes son rectas que intersectan el círculo en un único punto; como g manda líneas en líneas, manda entonces tangentes en tangentes. Como además g preserva paralelismo y puntos medios, tiene que dejar fijo el centro y por tanto es lineal.

Sea $\mathbf{u} = g(\boldsymbol{e}_1) \in \mathbb{S}^1$; nos falta demostrar que $g(\boldsymbol{e}_2) = \pm\mathbf{u}^{\perp}$. La tangente a $\mathbb{S}^1$ en $\boldsymbol{e}_1$, ℓ_1 digamos, es la vertical por $\boldsymbol{e}_1$, que bajo g tiene que ir a la tangente en $\mathbf{u}$, que es la ortogonal al vector $\mathbf{u}$ por el punto $\mathbf{u}$. Entonces, el eje y (paralelo a ℓ_1 por el centro) va bajo g a la ortogonal a $\mathbf{u}$ por el origen, sea ésta ℓ_2. Como $\boldsymbol{e}_2 \in \mathbb{S}^1 \cap (\text{eje } y)$, entonces $g(\boldsymbol{e}_2) \in \mathbb{S}^1 \cap \ell_2 = \{\mathbf{u}^{\perp}, -\mathbf{u}^{\perp}\}$, y por lo tanto g es ortogonal y se escribe como $g(\mathbf{x}) = \mathrm{B}\mathbf{x}$ con $\mathrm{B} \in \mathbf{O}(2)$. Tenemos entonces la siguiente igualdad de polinomios



$$\begin{aligned} \mathrm{P}_1(\mathbf{x}) &= k(\mathrm{P} \circ g)(\mathbf{x}) = k(\mathrm{P}(\mathrm{B}\mathbf{x})) \\ &= k\left(\mathbf{x} \cdot \left(\mathrm{B}^{\top}\mathrm{I}\mathrm{B}\right)\mathbf{x} - 1\right) = k(\mathbf{x} \cdot \mathrm{I}\mathbf{x} - 1) = k\,\mathrm{P}(\mathbf{x}), \end{aligned}$$

pues $\mathrm{B}^{\top}\mathrm{B} = \mathrm{I}$, que es lo que queríamos demostrar: *los polinomios cuadráticos que definen el círculo unitario,* $\mathbb{S}^1$, *son exactamente los de la forma*

$$k\left(x^2 + y^2 - 1\right) \quad \text{con} \quad k \neq 0.$$

Con este caso concreto en la mano, es fácil argumentar que para cualquier elipse los polinomios que la definen difieren por constantes distintas de cero. Pero en general, esta línea de argumentación deja mucho que desear. Habría que argumentar geométricamente en cada caso (llevamos uno de nueve) y obtener el grupo de simetrías afines de una cónica (véase el ejercicio sobre simetrías de la hipérbola). Algo que resulta imposible para las cónicas vacías, de las que tenemos dos, la degenerada y la no degenerada, que no podemos diferenciar geométricamente.

La línea de argumentación clásica va por otro lado; es algebraica. Pero depende de cerrar algebraicamente a los números reales en los complejos, y ahí sí funciona. Si dos polinomios tienen los mismos ceros, entonces, salvo constantes, son potencias de un mismo polinomio, y esto podría considerarse como el principio de la geometría algebraica moderna. Rebasa el contexto de este libro.

EJERCICIO 4.15 Demuestra que el conjunto de afinidades que dejan invariante un polinomio cuadrático P,

$$\mathbf{Sim}_{\mathbf{Af}}(\mathrm{P}) := \{g \in \mathbf{Af}(2) \mid \mathrm{P} \circ g = \mathrm{P}\},$$

es un grupo.

EJERCICIO 4.16 Para $\mathrm{P}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{x} - 1$ demuestra algebraicamente que

$$\mathbf{Sim}_{\mathbf{Af}}(\mathrm{P}) = \mathbf{O}(2).$$

EJERCICIO 4.17 Sea $\mathcal{H}$ la hipérbola canónica dada por la ecuación $x^2 - y^2 = 1$. Demuestra que para cualquier $(a, b) \in \mathcal{H}$ se tiene que la transformación lineal definida por la matriz

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix}$$

deja invariante a (es una simetría afín de) $\mathcal{H}$.

***EJERCICIO 4.18** Demuestra que si $g \in \mathbf{Af}(2)$ es tal que $g(\mathcal{H}) = \mathcal{H}$ (con $\mathcal{H}$ la hipérbola canónica) entonces g es lineal y está dada por una matriz como la anterior, o bien por esa seguida de la reflexión en el eje x.

***EJERCICIO 4.19** Demuestra que dos polinomios que definen la hipérbola canónica difieren por una constante ($P_1 = k\,P_2$).

Capítulo 5

La esfera y el espacio

En este capítulo estudiaremos la geometría de la esfera. Veremos cómo hay, y se puede hacer, geometría en otros "espacios" además del plano euclidiano. Pero antes tenemos que repasar y ver ciertas propiedades de su "hábitat" natural, del espacio ambiente en el que vive.

5.1 Planos y líneas en $\mathbb{R}^3$

Hemos visto que un plano en $\mathbb{R}^3$ se puede definir, al menos, de tres maneras. La primera, una *descripción paramétrica*, dando un punto base $\mathbf{p}$ y dos vectores direccionales no paralelos $\mathbf{u}$ y $\mathbf{v}$, para describir un plano Π con dos parámetros libres s y t como

$$\Pi = \{\, \mathbf{p} + s\,\mathbf{u} + t\,\mathbf{v} \mid s, t \in \mathbb{R} \,\}.$$

La segunda, una *descripción baricéntrica*, dando tres puntos no colineales $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ en él, y se le describe por coordenadas baricéntricas (o combinaciones afines):

$$\Pi = \{\, \alpha\,\mathbf{a} + \beta\,\mathbf{b} + \gamma\,\mathbf{c} \mid \alpha, \beta, \gamma \in \mathbb{R};\ \alpha + \beta + \gamma = 1 \,\}.$$

Y finalmente, una *descripción intrínseca* por medio de una *ecuación normal*: se dan un vector normal $\mathbf{n} \neq \mathbf{0}$ y una constante $d \in \mathbb{R}$ tales que

$$\Pi : \quad \mathbf{n} \cdot \mathbf{x} = d.$$

Pasar de una descripción baricéntrica a una paramétrica, y viceversa, es fácil pues tomando $\mathbf{a} = \mathbf{p}, \mathbf{b} = \mathbf{p} + \mathbf{u}$ y $\mathbf{c} = \mathbf{p} + \mathbf{v}$ por ejemplo, podemos transformar una en la otra (Sección 1.4.2). El primer problema importante que vamos a atacar es cómo encontrar una ecuación normal para un plano dado paramétricamente. Es decir, dados $\mathbf{u}$ y $\mathbf{v}$, no paralelos, encontrar un vector $\mathbf{n}$ perpendicular a ambos. Pero antes, veámos con un ejemplo cómo es fácil encontrar una descripción paramétrica, o baricéntrica, de un plano dado por una ecuación normal.

Tomemos por ejemplo la ecuación

$$2x + y - z = 2,$$

que define un plano Π normal al vector $(2, 1, -1)$. Para encontrar una descripción baricéntrica basta encontrar tres soluciones particulares no colineales. Y para esto, podemos asignar valores arbitrarios a dos de las variables y despejar la tercera. Lo más fácil es hacerlas cero. Así, si hacemos $y = 0$ y $z = 0$ la ecuación implica que $x = 1$ y por tanto que $\mathbf{a} = (1, 0, 0)$ es una solución. Análogamente, $\mathbf{b} = (0, 2, 0)$ y $\mathbf{c} = (0, 0, -2)$ son soluciones. Y estos tres puntos nos dan una descripción baricéntrica del mismo plano pues son los puntos de intersección de Π con los tres ejes, y en este caso no son colineales.

Para obtener una descripción paramétrica, podemos tomar a $\mathbf{a}$ como punto base y a $\mathbf{b} - \mathbf{a}$ y $\mathbf{c} - \mathbf{a}$ como vectores direccionales:

$$\begin{aligned} \Pi &= \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ -2 \end{pmatrix} \mid s, t \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} 1 - s - t \\ 2s \\ -2t \end{pmatrix} \mid s, t \in \mathbb{R} \right\}. \end{aligned}$$

Y efectivamente, estos valores siempre satisfacen la ecuación original; para corroborarlo, podemos sustituir en la ecuación original y ver que simbólicamente, es decir, independientemente de los valores de los parámetros, siempre se cumple que:

$$2\,(1 - s - t) + (2s) - (-2t) = 2.$$

Hay que observar que a veces puede ser un poco, pero no mucho, más difícil encontrar soluciones particulares no colineales (veánse los siguientes ejercicios).

Por último, es bonito ver lo fácil que resulta demostrar por qué el método que acabamos de esbozar funciona siempre.

Lema 5.1 *Si $\mathbf{a}, \mathbf{b}, \mathbf{c}$ son soluciones de la ecuación lineal $\mathbf{n} \cdot \mathbf{x} = d$, entonces cualquier combinación afín de ellos también lo es.*

Demostración. Recordemos que una combinación afín de $\mathbf{a}, \mathbf{b}, \mathbf{c}$ es un vector que se expresa de la forma

$$\alpha\, \mathbf{a} + \beta\, \mathbf{b} + \gamma\, \mathbf{c} \qquad \text{con} \qquad \alpha + \beta + \gamma = 1.$$

De la linearidad del producto interior, la hipótesis y esta última ecuación se obtiene:

$$\begin{aligned} \mathbf{n} \cdot (\alpha\, \mathbf{a} + \beta\, \mathbf{b} + \gamma\, \mathbf{c}) &= \alpha\, (\mathbf{n} \cdot \mathbf{a}) + \beta\, (\mathbf{n} \cdot \mathbf{b}) + \gamma\, (\mathbf{n} \cdot \mathbf{c}) \\ &= \alpha\, d + \beta\, d + \gamma\, d \\ &= (\alpha + \beta + \gamma)\, d = d. \end{aligned}$$

□

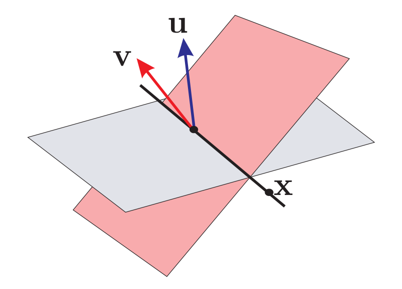
EJERCICIO 5.1 Encuentra descripciones baricéntricas y paramétricas para los siguientes planos: $\Pi_1 : 2x - 3y + z = 6$; $\Pi_2 : x + 4y - 2z = 4$; $\Pi_3 : -x + y - 2z = 2$; $\Pi_4 : 5x - 2y + z = 1$.

EJERCICIO 5.2 Enuncia y demuestra el lema anterior para combinaciones afines de longitud arbitraria (no necesariamente de tres vectores) y en cualquier espacio vectorial.

5.1.1 El producto cruz

Consideremos ahora el problema inverso: encontrar una ecuación normal para un plano descrito con parámetros. Puesto que sabemos pasar de una descripción baricéntrica a una paramétrica y viceversa, el problema se reduce a encontrar un vector perpendicular a dos vectores $\mathbf{u}$ y $\mathbf{v}$ dados. Es decir, tenemos que resolver el sistema

$$\begin{aligned} \mathbf{u} \cdot \mathbf{x} &= 0 \\ \mathbf{v} \cdot \mathbf{x} &= 0, \end{aligned}$$



o bien encontrar un punto $\mathbf{x}$ en la intersección de los planos ortogonales a $\mathbf{u}$ y a $\mathbf{v}$. Puesto que el sistema es *homogéneo* (las constantes son cero), siempre existe la solución *trivial* $\mathbf{x} = \mathbf{0}$; pero ésa no nos sirve pues el vector normal a un plano debe ser distinto de cero. De tal manera que si encontramos $\mathbf{n} \neq \mathbf{0}$ que satisfaga las dos ecuaciones anteriores, entonces el plano que pasa por $\mathbf{p}$ con vectores direccionales $\mathbf{u}$ y $\mathbf{v}$ tendrá la ecuación normal

$$\mathbf{n} \cdot \mathbf{x} = \mathbf{n} \cdot \mathbf{p}.$$

Además, resolvamos de una buena vez el problema de manera general, tratando de no hacer suposiciones adicionales sobre los vectores dados.

Asignemos coordenadas a nuestros vectores usando subíndices, $\mathbf{u} = (u_1, u_2, u_3)$ y $\mathbf{v} = (v_1, v_2, v_3)$, para que el sistema se escriba

$$\begin{aligned} u_1\, x + u_2\, y + u_3\, z &= 0 \\ v_1\, x + v_2\, y + v_3\, z &= 0 \end{aligned}$$

con $\mathbf{x} = (x, y, z)$ como nuestro vector variable o incógnita. Para eliminar una de las variables, podemos usar el mismo método que en sistemas de 2×2. Eliminemos z multiplicando la segunda ecuación por u_3, la primera por $-v_3$ y sumando para obtener

$$(u_3 v_1 - v_3 u_1)\, x + (u_3 v_2 - v_3 u_2)\, y = 0.$$

Esta ecuación ya la conocemos aunque las constantes tienen otros nombres (recuérdese nuestro primer encuentro con el compadre ortogonal). Tiene una solución canónica “volteando” los coeficientes y en uno de ellos también el signo, es decir:

$$x = u_2 v_3 - v_2 u_3 \ ; \ y = u_3 v_1 - v_3 u_1 .$$

Al sustituir ahora estos valores en la primera ecuación y simplificar se obtiene

$$\begin{aligned} u_1 \ (u_2 v_3 - v_2 u_3) + u_2 \ (u_3 v_1 - v_3 u_1) + u_3 \, z &= 0 \\ u_3 \, (u_2 v_1 - v_2 u_1 + z) &= 0, \end{aligned}$$

que, independientemente del valor de u_3, tiene la solución

$$z = u_1 v_2 - v_1 u_2 .$$

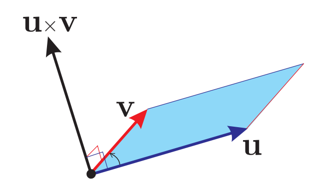
La solución particular que acabamos de encontrar (y que siempre funciona pues nunca dividimos ni hicimos ninguna suposición sobre si los coeficientes son cero o no) es tan importante que tiene nombre establecido. Se llama el *producto cruz* de $\mathbf{u}$ y $\mathbf{v}$. Y para resumir, remarcar y establecer su notación, definámoslo como el vector

$$\mathbf{u} \times \mathbf{v} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} := \begin{pmatrix} u_2 v_3 - v_2 u_3 \\ u_3 v_1 - v_3 u_1 \\ u_1 v_2 - v_1 u_2 \end{pmatrix} . \tag{5.1}$$

La fórmula parece complicada, pero la “cantaleta” que nos ayuda a recordarla la hace sencilla. Obsérvese que en la primera coordenada del producto cruz no aparecen las primeras coordenadas de los ingredientes: es justo el determinante de la matriz 2×2 que se obtiene al eliminar el primer renglón de la matriz 3×2 de los dos vectores. La segunda coordenada es de nuevo el determinante al eliminar ahora el segundo renglón, pero con signo negativo. Y la tercera vuelve a ser el determinante tal cual al olvidarse del tercer renglón. El producto cruz se arma de subdeterminantes 2×2 ignorando datos (coordenadas) correspondientes y teniendo cuidado en los signos (el de enmedio lleva un signo “menos”).

Geométricamente, el producto cruz es a $\mathbb{R}^3$ lo que el compadre ortogonal es a $\mathbb{R}^2$. En el plano, dado un vector $\mathbf{u}$, se puede escoger un vector canónico ortogonal a él girándolo 90° en la orientación positiva; ése es su compadre ortogonal $\mathbf{u}^\perp$ que tanto nos ha ayudado; nótese que $\mathbf{u}^\perp$ está en la media recta de vectores ortogonales a $\mathbf{u}$ que está a “su izquierda” y que tiene la magnitud de $\mathbf{u}$.

En $\mathbb{R}^3$ no es posible escoger de manera coherente un ortogonal para un vector variable $\mathbf{u}$, pues $\mathbf{u}$ tiene todo un plano de ortogonales y es imposible decidir cuál de ellos es "el bueno". Pero para dos vectores, $\mathbf{u}$ y $\mathbf{v}$, sí se puede elegir al "bueno" de manera geométrica: si generan un plano, tenemos que escoger un vector en su recta ortogonal. Primero escogemos el lado de ese plano donde el giro de $\mathbf{u}$ a $\mathbf{v}$ se ve en sentido contrario a las manesillas del reloj (positivo). Ello nos da media recta de sus posibles ortogonales. Y de estos, escogemos el que tiene magnitud igual al área del paralelogramo que generan $\mathbf{u}$ y $\mathbf{v}$. Así obtenemos un ortogonal a ellos, escogido canónicamente, y que es precisamente su producto cruz, $\mathbf{u} \times \mathbf{v}$, como veremos más adelante. Obsérvese además que la *construcción geométrica* del producto cruz que acabamos de esbozar funciona también cuando $\mathbf{u}$ y $\mathbf{v}$ no generan un plano, pues entonces están en una línea que pasa por el origen, el área de su paralelogramo es cero y les toca entonces el vector $\mathbf{0}$. Así que el producto cruz también debe detectar si $\mathbf{u}$ y $\mathbf{v}$ son linealmente dependientes.



Hemos hecho muchas afirmaciones que ahora debemos demostrar. Empecemos por las propiedades elementales del producto cruz con respecto a las otras operaciones que ya conocemos.

Teorema 5.2 *Sean* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ *tres vectores cualesquiera y* $t \in \mathbb{R}$. *Se cumplen entonces las siguientes propiedades del producto cruz:*

$$
\begin{array}{rl}
i) & (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{u} = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{v} = 0 \\
ii) & \mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u} \\
iii) & \mathbf{u} \times (\mathbf{v} + \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) + (\mathbf{u} \times \mathbf{w}) \\
iv) & \mathbf{u} \times (t\mathbf{v}) = t\,(\mathbf{u} \times \mathbf{v}) \\
v) & (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) \\
vi) & (\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\,\mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\,\mathbf{u} \\
vii) & |\mathbf{u} \times \mathbf{v}| = |\mathbf{u}|\,|\mathbf{v}| \operatorname{sen} \theta,
\end{array}
$$

donde θ *(entre* 0 *y* π*) es el ángulo entre* $\mathbf{u}$ *y* $\mathbf{v}$.

Demostración. La propiedad **(i)** es justo la motivación que usamos para definir el producto cruz (que es ortogonal a $\mathbf{u}$ y a $\mathbf{v}$). La propiedad **(ii)** se conoce como *antisimetría*: las variables no conmutan sino que *anticonmutan*. Las propiedades **(iii)** y **(iv)** son la *linearidad* del producto cruz, que junto con **(ii)** dan que también es lineal en la primera variable; se dice que es *bilineal*. Todas ellas se siguen directamente de la definición dada en (5.1) y se las dejamos como ejercicio al lector.

El inciso **(v)** es igualmente fácil pero es tan importante que preferimos darle un pequeño espacio. Al desarrolar ambos lados (según las definiciones de los productos cruz e interior, y usando la notación que más conviene, que es la de subíndices para las coordenadas: $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$ y $\mathbf{w} = (w_1, w_2, w_3)$) se obtiene una

suma de monomios de la forma $\pm u_i v_j w_k$ con $\{i, j, k\} = \{1, 2, 3\}$. A saber, siguiendo el orden de $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ y la correspondiente fórmula (5.1) para $(\mathbf{v} \times \mathbf{w})$, tenemos

$$\begin{aligned} & u_1 (v_2 w_3 - v_3 w_2) + u_2 (v_3 w_1 - v_1 w_3) + u_3 (v_1 w_2 - v_2 w_1) \\ = \; & u_1 v_2 w_3 + u_2 v_3 w_1 + u_3 v_1 w_2 - (u_1 v_3 w_2 + u_2 v_1 w_3 + u_3 v_2 w_1), \end{aligned}$$

que son todas las posibles permutaciones de los tres índices. Nótese que las que tienen signo positivo son las tres que preservan el orden cíclico $1 \to 2 \to 3 \to 1$ y las que tienen menos son las tres que lo invierten. Para concluir **(v)** basta ver que al desarrollar $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$ se obtienen de nuevo los seis posibles monomios y con los mismos signos, aunque en diferente orden. Esta función que asocia a tres vectores un "numerito" es el *determinante* y la estudiaremos con detalle más adelante.

La razón geométrica de **(vi)** es que cualquier vector ortogonal a $(\mathbf{u} \times \mathbf{v})$ está contenido en el plano que generan $\mathbf{u}$ y $\mathbf{v}$, y por tanto se expresa en términos de ellos. Pero que los coeficientes sean tan nítidos es agradable. Para demostrarlo hay que desarrollar

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = \begin{pmatrix} u_2 v_3 - v_2 u_3 \\ u_3 v_1 - v_3 u_1 \\ u_1 v_2 - v_1 u_2 \end{pmatrix} \times \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}.$$

Hagamos con cuidado la primera coordenada; las otras dos son análogas y se dejan al lector o bien se siguen de la simetría en el uso de los subíndices. Al aplicar de nuevo la fórmula del producto cruz se tiene que la primera coordenada es

$$\begin{aligned} & (u_3 v_1 - v_3 u_1)\, w_3 - (u_1 v_2 - v_1 u_2)\, w_2 \\ = \; & (u_3 w_3 + u_2 w_2)\, v_1 - (v_3 w_3 + v_2 w_2)\, u_1 \\ = \; & (u_1 w_1 + u_2 w_2 + u_3 w_3)\, v_1 - (v_1 w_1 + v_2 w_2 + v_3 w_3)\, u_1 \\ = \; & (\mathbf{u} \cdot \mathbf{w})\, v_1 - (\mathbf{v} \cdot \mathbf{w})\, u_1, \end{aligned}$$

que claramente es la primera coordenada de $(\mathbf{u} \cdot \mathbf{w})\, \mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\, \mathbf{u}$, y entonces

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\, \mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\, \mathbf{u}.$$

Probemos, antes de entrarle a **(vii)**, que de aquí se sigue que el producto cruz **no es asociativo** (más que un "producto algebraico" es un "producto geométrico"). Usando la antisimetría, **(ii)**, y la fórmula para un doble producto, **(vi)**, se obtiene

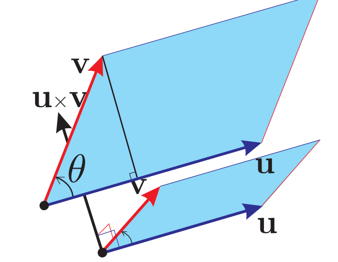
$$\begin{aligned} \mathbf{u} \times (\mathbf{v} \times \mathbf{w}) &= -(\mathbf{v} \times \mathbf{w}) \times \mathbf{u} \\ &= -((\mathbf{v} \cdot \mathbf{u})\, \mathbf{w} - (\mathbf{w} \cdot \mathbf{u})\, \mathbf{v}) \\ &= (\mathbf{u} \cdot \mathbf{w})\, \mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\, \mathbf{w}, \end{aligned}$$

que no tiene por qué ser igual a $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$. Geométricamente, este último está en el plano $\langle \mathbf{u}, \mathbf{v} \rangle$ y el que desarrollamos en el $\langle \mathbf{v}, \mathbf{w} \rangle$.

Y para demostrar **(vii)** desarrollemos $|\mathbf{u}\times\mathbf{v}|^2$ usando, respectivamente, **(v)** y **(vi)** en las dos primeras igualdades:

$$\begin{aligned}(\mathbf{u}\times\mathbf{v})\cdot(\mathbf{u}\times\mathbf{v}) &= \mathbf{u}\cdot(\mathbf{v}\times(\mathbf{u}\times\mathbf{v}))\\ &= \mathbf{u}\cdot((\mathbf{v}\cdot\mathbf{v})\,\mathbf{u}-(\mathbf{v}\cdot\mathbf{u})\,\mathbf{v})\\ &= (\mathbf{v}\cdot\mathbf{v})\,(\mathbf{u}\cdot\mathbf{u})-(\mathbf{u}\cdot\mathbf{v})^2\\ &= |\mathbf{v}|^2\,|\mathbf{u}|^2-|\mathbf{u}|^2\,|\mathbf{v}|^2\cos^2\theta\\ &= |\mathbf{u}|^2\,|\mathbf{v}|^2\left(1-\cos^2\theta\right)=|\mathbf{u}|^2\,|\mathbf{v}|^2\,\mathrm{sen}^2\,\theta.\end{aligned}$$

De donde se sigue **(vii)**, pues todos los factores son positivos. □



Corolario 5.3 *La magnitud de* $\mathbf{u}\times\mathbf{v}$ *es el área del paralelogramo generado por* $\mathbf{u}$ *y* $\mathbf{v}$*.*

Demostración. Si tomamos a $\mathbf{u}$ como base del paralelogramo, entonces su altura es $|\mathbf{v}|\,\mathrm{sen}\,\theta$; y el corolario se sigue de **(vii)**. □

Corolario 5.4 *Los vectores* $\mathbf{u}$ *y* $\mathbf{v}$ *son paralelos si y sólo si su producto cruz,* $\mathbf{u}\times\mathbf{v}$*, es igual a* $\mathbf{0}$*.*

Demostración. Esto se sigue de **(vii)**, pues $\mathbf{u}$ y $\mathbf{v}$ (no nulos) son paralelos si y sólo si el ángulo θ entre ellos es 0 o π, y esto sucede si y sólo si $\mathrm{sen}\,\theta=0$. □

Nótese además que si $\mathbf{u}$ o $\mathbf{v}$ son el vector cero, entonces su producto cruz también es cero. Por lo tanto tenemos que $\mathbf{u}$ y $\mathbf{v}$ generan un plano si y sólo si su producto cruz es distinto de cero, y en este caso $\mathbf{u}\times\mathbf{v}$ es normal a ese plano.

EJERCICIO 5.3 Encuentra una ecuación normal para los siguientes planos:
$\Pi_1=\{(2+t-s\,,\,1-2t+s\,,\,3t-s-3)\mid t,s\in\mathbb{R}\}$
$\Pi_2=\{(\alpha+\beta-2\gamma\,,\,2\alpha-2\beta+\gamma\,,\,3\alpha+\beta-\gamma)\mid \alpha,\beta,\gamma\in\mathbb{R}\,;\ \alpha+\beta+\gamma=1\}$
$\Pi_3=\{(2\alpha+\beta-\gamma\,,\,\alpha-3\beta+\gamma\,,\,\alpha-\beta+\gamma)\mid \alpha,\beta,\gamma\in\mathbb{R}\,;\ \alpha+\beta+\gamma=1\}$
$\Pi_4=\{(3+2t+s\,,\,2s-t\,,\,t+5s-3)\mid t,s\in\mathbb{R}\}$

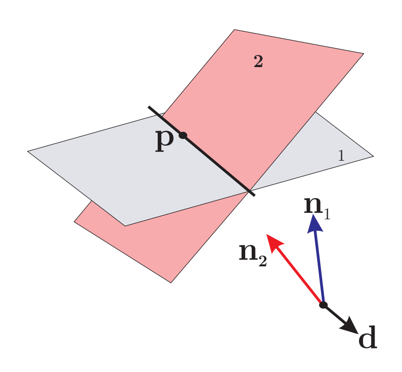
EJERCICIO 5.4 ¿Cuáles de las siguientes cuartetas de puntos en $\mathbb{R}^3$ son coplanares? Y si lo son define el plano en el que están.
$\mathbf{a}=(2,-1,0)\,,\mathbf{b}=(1,2,-2)\,,\mathbf{c}=(0,1,0)\,,\mathbf{d}=(4,-1,-1)\,.$
$\mathbf{a}=(2,-1,0)\,,\mathbf{b}=(1,2,-2)\,,\mathbf{c}=(0,1,0)\,,\mathbf{d}=(3,-1,-1)\,.$

EJERCICIO 5.5 Encuentra un criterio general para saber si cuatro puntos $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}\in\mathbb{R}^3$ son coplanares o no, y demuéstralo.

5.1.2 Intersección de planos

Consideremos dos planos dados por una ecuación normal

$$\begin{aligned} \Pi_1 &: \mathbf{n}_1 \cdot \mathbf{x} = d_1 \\ \Pi_2 &: \mathbf{n}_2 \cdot \mathbf{x} = d_2 . \end{aligned} \tag{5.2}$$



Supongamos que sus dos vectores normales no son paralelos. Entonces el vector $\mathbf{d} := \mathbf{n}_1 \times \mathbf{n}_2$ es no nulo por el Corolario 5.4. Ésta es la única "dirección común" a los dos planos, pues es normal a los dos vectores normales; de tal manera que la recta donde se intersectan debe tener esa dirección. Supongamos que $\mathbf{p}$ es una solución particular al sistema de dos ecuaciones con tres incógnitas anterior, es decir, que $\mathbf{p}$ es un punto en la intersección: $\mathbf{p} \in \Pi_1 \cap \Pi_2$. Vamos a demostrar que entonces la recta

$$\ell = \{\mathbf{p} + t\,\mathbf{d} \mid t \in \mathbb{R}\}$$

es precisamente la intersección.

Primero veamos que un punto en ℓ satisface las ecuaciones (5.2). Para $i = 1, 2$ se tiene:

$$\begin{aligned} \mathbf{n}_i \cdot (\mathbf{p} + t\,\mathbf{d}) &= \mathbf{n}_i \cdot \mathbf{p} + \mathbf{n}_i \cdot (t\,\mathbf{d}) \\ &= \mathbf{n}_i \cdot \mathbf{p} + t\,(\mathbf{n}_i \cdot \mathbf{d}) = \mathbf{n}_i \cdot \mathbf{p} = d_i \end{aligned}$$

pues $\mathbf{n}_i \cdot \mathbf{d} = 0$ por definición de $\mathbf{d}$. Lo que demuestra que $\ell \subset \Pi_1 \cap \Pi_2$. Y por el otro lado, si $\mathbf{q} \in \Pi_1 \cap \Pi_2$ probaremos que $\mathbf{q} - \mathbf{p}$ es paralelo a $\mathbf{d}$, para lo cual usamos el criterio del Corolario 5.4 y el Teorema 5.2:

$$\begin{aligned} \mathbf{d} \times (\mathbf{q} - \mathbf{p}) &= \mathbf{d} \times \mathbf{q} - \mathbf{d} \times \mathbf{p} \\ &= (\mathbf{n}_1 \times \mathbf{n}_2) \times \mathbf{q} - (\mathbf{n}_1 \times \mathbf{n}_2) \times \mathbf{p} \\ &= (\mathbf{n}_1 \cdot \mathbf{q})\,\mathbf{n}_2 - (\mathbf{n}_2 \cdot \mathbf{q})\,\mathbf{n}_1 - (\mathbf{n}_1 \cdot \mathbf{p})\,\mathbf{n}_2 + (\mathbf{n}_2 \cdot \mathbf{p})\,\mathbf{n}_1 \\ &= d_1\,\mathbf{n}_2 - d_2\,\mathbf{n}_1 - d_1\,\mathbf{n}_2 + d_2\,\mathbf{n}_1 = 0. \end{aligned}$$

Hemos completado la parte central de la demostración del siguiente teorema.

Teorema 5.5 *Dos planos en $\mathbb{R}^3$ se intersectan en una recta si y solo si sus vectores normales no son paralelos. En caso contrario, o no se intersectan (son* paralelos*) o son el mismo plano.*

Demostración. Nos falta nada más afinar detalles. Primero, suponiendo que los vectores normales no son paralelos, tenemos que ver que existe una solución particular $\mathbf{p}$ como la que usamos arriba; y segundo, observar que cuando son paralelos, que exista o no esta solución particular determina los dos casos correspondientes. Para esto, consideremos al sistema (5.2) en coordenadas:

$$\begin{aligned} a_1\,x + b_1\,y + c_1\,z &= d_1 \\ a_2\,x + b_2\,y + c_2\,z &= d_2 . \end{aligned}$$

Para encontrar una solución particular podemos asignar un valor arbitrario a cualquiera de las variables y resolver el sistema 2×2 que queda. Lo más fácil es hacerla 0; por ejemplo si $z = 0$ tenemos un sistema 2×2 clásico (en x y y) que tiene solución única dependiendo de su determinante. Puesto que las coordenadas del producto cruz son precisamente ($\pm$) estos determinantes, que los vectores normales no sean paralelos implica que alguno de los tres subsistemas 2×2 que se obtienen al hacer una variable 0 tiene determinante no nulo y por tanto tiene solución. En el caso en que son paralelos los vectores normales, entonces los lados izquierdos son múltiplos: si las constantes lo son por el mismo factor, definen al mismo plano, pero si no lo son no existen soluciones comunes. □

Asi que las rectas en $\mathbb{R}^3$ se pueden determinar por dos ecuaciones lineales o bien describirse paramétricamente.

EJERCICIO 5.6 Encuentra una descripción paramétrica para la recta de intersección de las siguientes parejas de planos:
$\Pi_1 : 2x + y - z = 1; \quad \Pi_2 : -2x + y - 3z = 3.$
$\Pi_1 : x - y - z = 0; \quad \Pi_2 : x + y - z = 1.$
$\Pi_1 : 2x + y - z = 2; \quad \Pi_2 : -x + y - 2z = 2.$
$\Pi_1 : 2x + z = 1; \quad \Pi_2 : -2x + z = 3.$

EJERCICIO 5.7 Describe las siguientes rectas *intrínsecamente*, es decir, como las soluciones de dos ecuaciones lineales:
$\ell_1 = \{(2 + t\, ,\, 1 - 2t\, ,\, 3t - 3) \mid t \in \mathbb{R}\}$
$\ell_2 = \{(s\, ,\, 2 - 3s\, ,\, 2s - 3) \mid s \in \mathbb{R}\}.$

EJERCICIO 5.8 Sea Π el plano dado por la ecuación $\mathbf{n} \cdot \mathbf{x} = c$ y sea ℓ la recta $\{\mathbf{p} + t\,\mathbf{d} \mid t \in \mathbb{R}\}$. Demuestra (sustituyendo la expresión de los puntos de ℓ en la ecuación de Π) que Π y ℓ se intersectan en un único punto si y sólo si $\mathbf{n} \cdot \mathbf{d} \neq 0$. Observa que si no es así (es decir, si $\mathbf{n} \cdot \mathbf{d} = 0$) entonces la dirección $\mathbf{d}$ es paralela al plano; por tanto demostraste que un plano y una recta se intersectan en un único punto si y sólo si la direccion de la recta no es paralela al plano.

EJERCICIO 5.9 Sean $\mathbf{u}, \mathbf{v}, \mathbf{w}$ tres vectores tales que $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} \neq 0$. Demuestra que tres planos normales a $\mathbf{u}, \mathbf{v}$ y $\mathbf{w}$ respectivamente se intersectan en un único punto.

5.1.3 El determinante y la orientación

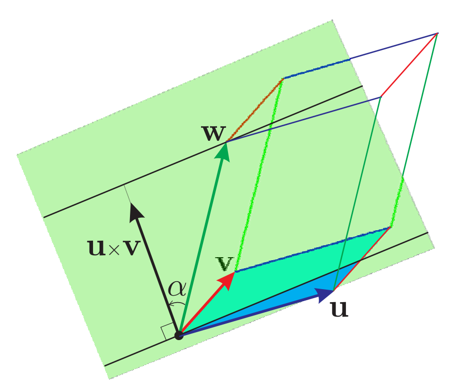
Dados tres vectores $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, definimos su *determinante* como el número real

$$\begin{aligned} \det(\mathbf{u}, \mathbf{v}, \mathbf{w}) &= (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} \\ &= (\mathbf{v} \times \mathbf{w}) \cdot \mathbf{u} \\ &= (\mathbf{w} \times \mathbf{u}) \cdot \mathbf{v}, \end{aligned} \tag{5.3}$$

donde hemos usado la propiedad **(v)** del producto cruz y la simetría del producto punto. El determinante tiene el siguiente significado geométrico.

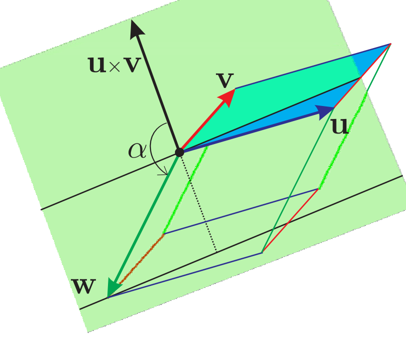
Proposicion 5.6 *Dados tres vectores* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$*, su determinante es el volumen signado del paralelepípedo que definen, y el signo corresponde a su "orientación".*

Demostración. Consideremos la fórmula $\det(\mathbf{u}, \mathbf{v}, \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$. Sabemos que $|\mathbf{u} \times \mathbf{v}|$ es el área del paralelogramo $\mathbf{u}, \mathbf{v}$ que podemos tomar como base del paralelepípedo en cuestión. Y por la fórmula geométrica del producto punto tenemos entonces que



$$\det(\mathbf{u}, \mathbf{v}, \mathbf{w}) = |\mathbf{u} \times \mathbf{v}|\ |\mathbf{w}| \cos \alpha,$$

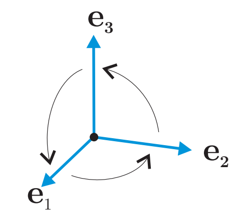
donde α es el ángulo entre $\mathbf{w}$ y $\mathbf{u} \times \mathbf{v}$.



Supongamos primero que $0 \leq \alpha \leq \pi/2$. Entonces, en el plano generado por $\mathbf{w}$ y $\mathbf{u} \times \mathbf{v}$ (en el que el plano $\langle \mathbf{u}, \mathbf{v} \rangle$ —generado por $\mathbf{u}$ y $\mathbf{v}$, recuérdese— se ve como la línea perpendicular a $\mathbf{u} \times \mathbf{v}$), es fácil ver que la altura del paralelepípedo es $|\mathbf{w}| \cos \alpha$ y se sigue el resultado: tenemos que $\det(\mathbf{u}, \mathbf{v}, \mathbf{w})$ es "base por altura". Si $\pi/2 \leq \alpha \leq \pi$ entonces $|\mathbf{w}| \cos \alpha$ es también esa altura pero con signo negativo. Corresponde a que $\mathbf{w}$ vive en el otro lado del plano $\langle \mathbf{u}, \mathbf{v} \rangle$ en el que está $\mathbf{u} \times \mathbf{v}$. Pero en valor absoluto, $\det(\mathbf{u}, \mathbf{v}, \mathbf{w})$ sigue siendo el volumen del paralelepípedo que definen $\mathbf{u}, \mathbf{v}, \mathbf{w}$. □

El producto cruz $\mathbf{u} \times \mathbf{v}$ "escoge" un lado *positivo* del plano $\langle \mathbf{u}, \mathbf{v} \rangle$; los vectores $\mathbf{w}$ que están en ese lado (un *semiespacio*) ven el giro de $\mathbf{u}$ a $\mathbf{v}$ como la orientación positiva del plano $\mathbb{R}^2$ (en contra de las manecillas del reloj). Por convención en cómo se coloca la base canónica, el lado positivo corresponde a la regla de la mano derecha: si con la punta de los dedos de la mano derecha simulamos el giro de $\mathbf{u}$ a $\mathbf{v}$, entonces el pulgar apunta hacia el lado positivo del plano, hacia $\mathbf{u} \times \mathbf{v}$. Para los vectores $\mathbf{w}$ que están en el otro lado, el *negativo*, decimos que la terna $\mathbf{u}, \mathbf{v}, \mathbf{w}$ induce la *orientación negativa* a $\mathbb{R}^3$.

Así, por ejemplo, la terna canónica $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ en su orden canónico tiene orientación positiva, pero la terna $\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3$ tiene orientación negativa. De tal manera que si ponemos las coordenadas de la pantalla de una computadora con el eje x ($\mathbf{e}_1$) horizontal a la derecha (como se lee), y el segundo eje (y o el vector $\mathbf{e}_2$) hacia arriba; entonces el lado positivo del eje z (o el tercer vector canónico $\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2$) apunta hacia nosotros. Pero si decidiéramos que el lado positivo del eje z vaya hacia atrás, hacia el otro lado de la pantalla, y disponemos los vectores de acuerdo con estas coordenadas, entonces las ternas con determinante positivo corresponden a las que cumplen la regla de la mano izquierda, pues la base canónica lo hace y se cumple que $\det(\mathrm{I}) = 1 > 0$. En fin, es una convención que depende de cómo se situe la base canónica; aquí podemos suponer que es como en la figura, con la regla de la mano derecha.



Proposicion 5.7 *Sean* $\mathbf{u}$ *y* $\mathbf{v}$ *dos vectores que generan un plano* $\langle \mathbf{u}, \mathbf{v} \rangle$ *en* $\mathbb{R}^3$. *Entonces* $\mathbf{w}$ *está en el plano* $\langle \mathbf{u}, \mathbf{v} \rangle$ *si y sólo si* $\det(\mathbf{u}, \mathbf{v}, \mathbf{w}) = 0$.

Demostración. Por el Corolario 5.4 tenemos que $\mathbf{u} \times \mathbf{v} \neq \mathbf{0}$. Entonces el plano definido por la ecuación $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{x} = 0$ es precisamente el plano generado por $\mathbf{u}$ y $\mathbf{v}$, y se sigue la proposición, pues $\det(\mathbf{u}, \mathbf{v}, \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$. □

En términos de la Proposición anterior, $\mathbf{w}$ está en el plano $\langle \mathbf{u}, \mathbf{v} \rangle$ si y sólo si el volumen del paralelepípedo definido por $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ es cero.

EJERCICIO 5.10 Demuestra que el determinante cumple las siguientes propiedades

i) $\det(\mathbf{u}, \mathbf{v}, \mathbf{w}) = -\det(\mathbf{v}, \mathbf{u}, \mathbf{w}) = -\det(\mathbf{u}, \mathbf{w}, \mathbf{v}) = -\det(\mathbf{w}, \mathbf{v}, \mathbf{u})$
ii) $\det(\mathbf{u}, \mathbf{u}, \mathbf{v}) = 0$
iii) $\det(t\,\mathbf{u}, \mathbf{v}, \mathbf{w}) = t\,\det(\mathbf{u}, \mathbf{v}, \mathbf{w})$
iv) $\det(\mathbf{u} + \mathbf{x}, \mathbf{v}, \mathbf{w}) = \det(\mathbf{u}, \mathbf{v}, \mathbf{w}) + \det(\mathbf{x}, \mathbf{v}, \mathbf{w})$

EJERCICIO 5.11 Demuestra, usando únicamente el ejercicio anterior, que el determinante no cambia si sumamos un múltiplo de un vector a alguno de los otros, es decir, que

$$\det(\mathbf{u}, \mathbf{v} + t\,\mathbf{u}, \mathbf{w}) = \det(\mathbf{u}, \mathbf{v}, \mathbf{w}).$$

5.1.4 Sistemas de ecuaciones

Aplicaremos ahora el producto cruz a la resolución de sistemas de tres ecuaciones lineales con tres incógnitas. Como vimos en el Capítulo 3, tal sistema se puede escribir como

$$A\mathbf{x} = \mathbf{b},$$

donde A es una matriz de 3×3, $\mathbf{x}$ es el vector variable, o de las incógnitas (x, y, z), y $\mathbf{b} \in \mathbb{R}^3$ es un vector constante. Si denotamos por $\mathbf{u}, \mathbf{v}, \mathbf{w}$ a las columnas de A (es decir, si $A = (\mathbf{u}, \mathbf{v}, \mathbf{w})$) entonces es natural definir el *determinante de la matriz* A como

$$\det A = \det(\mathbf{u}, \mathbf{v}, \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$$

y tenemos:

Teorema 5.8 *El sistema* $A\mathbf{x} = \mathbf{b}$ *tiene solución única si y sólo si* $\det A \neq 0$.

Demostración. Si pensamos el sistema de ecuaciones $A\mathbf{x} = \mathbf{b}$ como tres ecuaciones lineales y por lo tanto, si son no triviales, como tres planos, la solución del sistema es la intersección de los tres planos. Por la Sección 5.1.2 y los Ejercicios 5.8 y 5.9 se tiene una buena pàrte del Teorema. Pero nótese que entonces los vectores importantes, los normales a los planos, corresponden a los renglones de A (digamos que son $\mathbf{n}_1$, $\mathbf{n}_2$, $\mathbf{n}_3$) y habría que revisar que el determinante se calcula igual ($\det A = (\mathbf{n}_1 \times \mathbf{n}_2) \cdot \mathbf{n}_3$, es decir, $\det A = \det A^\top$). De alguna manera, ya hicimos parte del trabajo. Pero veamos otra demostración pensando en la matriz A como tres vectores columna.

Con la notación del párrafo anterior ($A = (\mathbf{u}, \mathbf{v}, \mathbf{w})$, y $\mathbf{x} = (x, y, z)^\top$ el vector variable o incógnita), el sistema $A\mathbf{x} = \mathbf{b}$ es equivalente a la ecuación vectorial

$$x\,\mathbf{u} + y\,\mathbf{v} + z\,\mathbf{w} = \mathbf{b},$$

donde todos los vectores están dados. Tomando el producto punto con el producto cruz de dos de los vectores se eliminan las dos variables correspondientes. Así, para eliminar x y y de un solo golpe, tomamos el producto punto con $(\mathbf{u} \times \mathbf{v})$ para obtener la ecuación

$$z\ ((\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{b},$$

pues $(\mathbf{u} \times \mathbf{v})$ es ortogonal a $\mathbf{u}$ y a $\mathbf{v}$. De aquí, si $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \det A \neq 0$, se tiene que

$$z = \frac{(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{b}}{\det A}.$$

Y análogamente, por (5.3), $\det A \neq 0$ implica que

$$y = \frac{(\mathbf{w} \times \mathbf{u}) \cdot \mathbf{b}}{\det A}, \quad x = \frac{(\mathbf{v} \times \mathbf{w}) \cdot \mathbf{b}}{\det A}.$$

Por lo tanto, si $\det A \neq 0$ entonces la solución de $A\mathbf{x} = \mathbf{b}$ es única y está dada por las fórmulas anteriores que son conocidas como *"la regla de Krammer"*.

Ahora tenemos que demostrar el inverso. Supongamos entonces que $\det A = 0$ y hay que ver que el sistema no tiene solución única. Expresemos el determinante como $\det A = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$. Si $(\mathbf{u} \times \mathbf{v}) = 0$ entonces por el Corolario 5.4 tenemos que $\mathbf{u}$ y $\mathbf{v}$ son paralelos, y si $(\mathbf{u} \times \mathbf{v}) \neq 0$, entonces por la Proposición 5.7 tenemos que

$w \in \langle \mathbf{u}, v \rangle$. En ambos casos podemos afirmar que existen números x_0, y_0 y z_0 con alguno de ellos distinto de cero que cumplen

$$x_0\, \mathbf{u} + y_0\, v + z_0\, w = 0, \tag{5.4}$$

pues en el primer caso ($\mathbf{u} \times v = 0$) tenemos una igualdad $\mathbf{u} = s\,v$ o $v = r\,\mathbf{u}$, y en el segundo ($\mathbf{u} \times v \neq 0$) podemos expresar $w = r\,\mathbf{u} + s\,v$; y de cualquiera de estas igualdades obtenemos una expresión (5.4) con alguno de los coeficientes igual a 1 al pasar todos los sumandos al lado izquierdo.

Esto implica que el sistema homogéneo $A\mathbf{x} = \mathbf{0}$ no tiene solución única pues encontramos una diferente de la *trivial* que es $\mathbf{x} = \mathbf{0}$. Resumiendo, si suponemos que $\det A = 0$, hemos demostrado que existe $\mathbf{x}_0 = (x_0, y_0, z_0) \neq \mathbf{0}$ tal que $A\mathbf{x}_0 = \mathbf{0}$. Ésta será la base para demostrar que para cualquier $\mathbf{b} \in \mathbb{R}^3$ la solución de $A\mathbf{x} = \mathbf{b}$ no es única: ...Dibujo

Si el sistema $A\mathbf{x} = \mathbf{b}$ no tiene solución (el caso, por ejemplo, en el que w está en el plano generado por $\mathbf{u}$ y v pero $\mathbf{b}$ no está en ese plano) entonces, en particular, **no** hay solución única. Pero si $A\mathbf{x} = \mathbf{b}$ sí tiene solución, digamos que $\mathbf{x}_1$ es tal que $A\mathbf{x}_1 = \mathbf{b}$, entonces $\mathbf{x}_1 + \mathbf{x}_0$ también es solución, pues

$$A\,(\mathbf{x}_1 + \mathbf{x}_0) = A\mathbf{x}_1 + A\mathbf{x}_0 = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

Y como $\mathbf{x}_1 \neq \mathbf{x}_1 + \mathbf{x}_0$ pues $\mathbf{x}_0 \neq \mathbf{0}$ entonces la solución **no** es única. Dicho de otra manera, si existen x_1, y_1 y z_1 tales que

$$x_1\, \mathbf{u} + y_1\, v + z_1\, w = \mathbf{b},$$

entonces también se cumple que

$$(x_1 + x_0)\, \mathbf{u} + (y_1 + y_0)\, v + (z_1 + z_0)\, w = \mathbf{b}.$$

Lo cual concluye la demostración del teorema pues que no haya solución única quiere decir que o bien no hay solución o si sí la hay, entonces hay muchas (en el fondo hemos demostrado que cuando las hay, hay tantas soluciones como las hay para el sistema homogéneo $A\mathbf{x} = \mathbf{0}$). □

EJERCICIO 5.12 Sea $A = (\mathbf{u}, v, w)$ una matriz de 3×3 con $\det A \neq 0$. Demuestra que entonces tiene matriz inversa A^{-1} y que ésta está dada por la fórmula

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} (v \times w)^{\top} \\ (w \times \mathbf{u})^{\top} \\ (\mathbf{u} \times v)^{\top} \end{pmatrix},$$

donde, recuerda, $\mathbf{x}^{\top}$ es el vector renglón transpuesto del vector columna $\mathbf{x}$.

5.1.5 Dependencia e independencia lineal

En la demostración del regreso (si $\det A = 0$, la solución no es única) del teorema anterior hubo un paso fundamental que amerita comentarios y profundización. El de encontrar una solución no trivial a la ecuación homogénea $A\mathbf{x} = \mathbf{0}$, porque es la manera más democrática y general de decir que las columnas de la matriz A no son tan generales como podrían ser; que en vez de generar todo $\mathbb{R}^3$ sólo generan un plano o una línea o el origen (si las tres son $\mathbf{0}$). Amerita una pomposa definición que funciona en general, aunque en este libro sólo se le dé uso en instancias pequeñas ($k \leq 4$ y $n \leq 3$).

Definición 5.1.1 Sean $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k$ vectores en $\mathbb{R}^n$. Se dice que son *linealmente independientes* si

$$\sum_{i=1}^{k} \lambda_i \mathbf{u}_i = \mathbf{0} \quad \Rightarrow \quad \lambda_1 = \lambda_2 = \cdots = \lambda_k = 0,$$

es decir, si la única combinación lineal de ellos que da el vector $\mathbf{0}$ es la *trivial.* Y si, por el contrario, existen $\lambda_1, \lambda_2, \cdots, \lambda_k$ no todas ellas cero, tales que $\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \cdots + \lambda_k \mathbf{u}_k = \mathbf{0}$, entonces se dice que son *linealmente dependientes.*

Así se tiene que un vector es linealmente dependiente si y sólo si es el vector $\mathbf{0}$; que dos vectores no nulos son linealmente dependientes si y sólo si son paralelos, y que tres vectores son linealmente dependientes si y sólo si los tres están en un mismo plano por el origen. Estos son los casos que nos interesan y han quedado debida y elegantemente incluidos en una sola definición. Podemos decir entonces que el paso fundamental de la demostración anterior fue ver que $\det A = 0$ si y sólo si las columnas de A son linealmente dependientes. A estas matrices (con determinante cero) se les llama matrices *singulares*, y correspondientemente una matriz A es *no singular* si $\det A \neq 0$.

A manera de resumen, enlistemos una serie de equivalencias, cuya demostración o bien ya dimos o bien falta sólo algún argumento sencillo que dejamos como ejercicio mental para el lector.

Teorema 5.9 *Sea* $A = (\mathbf{u}, \mathbf{v}, \mathbf{w})$ *una matriz* 3×3*, y sea* $f : \mathbb{R}^3 \to \mathbb{R}^3$ *su función lineal asociada (es decir, definida por* $f(\mathbf{x}) = A\mathbf{x}$ *). Entonces son equivalentes:*

i) $\det \mathbf{A} \neq \mathbf{0}$ *(A es no singular)*
ii) $\mathbf{u}, \mathbf{v}, \mathbf{w}$ *son linealmente independientes*
iii) el sistema $\mathbf{Ax} = \mathbf{b}$ *tiene solución única para alguna* $\mathbf{b} \in \mathbf{R}^3$
iv) el sistema $\mathbf{Ax} = \mathbf{b}$ *tiene solución única para toda* $\mathbf{b} \in \mathbf{R}^3$
v) $\langle \mathbf{u}, \mathbf{v}, \mathbf{w} \rangle = \mathbf{R}^3$ *(*i.e., $\mathbf{u}, \mathbf{v}, \mathbf{w}$ *generan* $\mathbf{R}^3$)
vi) $\mathbf{f}$ *es inyectiva*
vii) $\mathbf{f}$ *es suprayectiva*
viii) $\mathbf{f}$ *es biyectiva*
ix) A *es* invertible *(existe* A^{-1} *tal que* $AA^{-1} = I$*).*

EJERCICIO 5.13 Demuestra que tres planos en $\mathbb{R}^3$ se intersectan en un único punto si y sólo si sus tres vectores normales son linealmente independientes.

EJERCICIO 5.14 Sean $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ linealmente independientes, y sean ℓ_1 y ℓ_2 dos rectas cualesquiera con direcciones $\mathbf{d}_1$ y $\mathbf{d}_2$ respectivamente. Demuestra que existe una única recta ℓ_3 con dirección $\mathbf{d}_3$ que pasa por ℓ_1 y ℓ_2.

...Dibujo

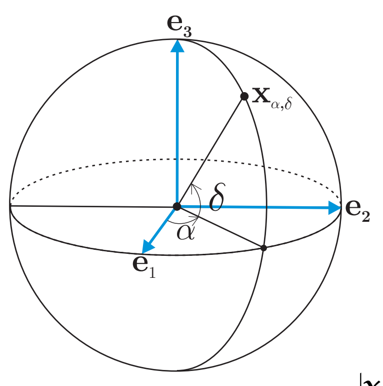
5.2 La esfera

Ahora sí, estudiaremos algo de la geometría de la *esfera*:

$$\mathbb{S}^2 := \left\{ \mathbf{x} \in \mathbb{R}^3 \mid |\mathbf{x}| = 1 \right\},$$

es decir, el conjunto de todos los vectores *unitarios* (de norma 1) en $\mathbb{R}^3$. No está de más remarcar que es la frontera de la "*bola sólida*" $\mathbb{B}^3 := \left\{ \mathbf{x} \in \mathbb{R}^3 \mid |\mathbf{x}| \leq 1 \right\}$ y que si bien esta última tiene tres dimensiones, su frontera, la esfera, tiene sólo dos; así que el exponente se refiere a la dimensión (se lee "ese dos"). Y conviene esta notación pues es parte de una familia, la *esfera de dimensión* n vive en el espacio euclidiano de la dimensión siguiente, $\mathbb{R}^{n+1}$, y ahí se define como el conjunto de vectores de norma 1. De esta familia ya estudiamos el círculo unitario $\mathbb{S}^1$, pero también hay una esfera de dimensión 3 y es interesante irse preguntando qué cosas valen ahí, aunque aquí sólo nos referiremos a la de dimensión 2. Intrínsecamente $\mathbb{S}^2$ es una superficie y su símil más parecido en el mundo real son las pompas de jabón.

Por su parte, la esfera es la representante teórica de cosas importantes como la superficie de la Tierra o el firmamento. Siguiendo este símil, podemos dar dos coordenadas angulares para sus puntos. El *ecuador* consta de los puntos



$$\{(\cos\alpha, \operatorname{sen}\alpha, 0) \mid \alpha \in [0, 2\pi]\},$$

donde α es llamada la *longitud*, y se identifica naturalmente con el círculo unitario $\mathbb{S}^1$. Y luego, para una longitud α dada podemos variar la *altitud* $\delta \in [-\pi, \pi]$ para obtener el punto

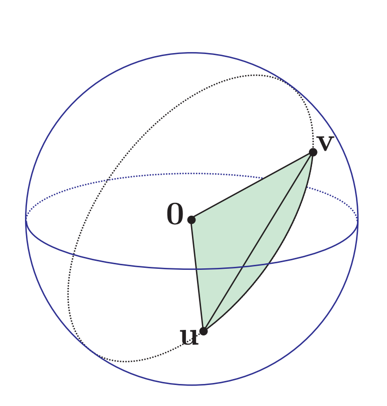
$$\mathbf{x}_{\alpha,\delta} = (\cos\delta\cos\alpha, \cos\delta\operatorname{sen}\alpha, \operatorname{sen}\delta),$$

que efectivamente está en la esfera pues

$$\begin{aligned} |\mathbf{x}_{\alpha,\delta}|^2 &= (\cos\delta\cos\alpha)^2 + (\cos\delta\operatorname{sen}\alpha)^2 + \operatorname{sen}^2\delta \\ &= \cos^2\delta\left(\cos^2\alpha + \operatorname{sen}^2\alpha\right) + \operatorname{sen}^2\delta \\ &= \cos^2\delta + \operatorname{sen}^2\delta = 1. \end{aligned}$$

Así que podemos llegar a cualquier punto de la esfera por sus dos coordenadas, longitud α y altitud δ, pero para el "polo norte", $\boldsymbol{e}_3$, y el "polo sur", $-\boldsymbol{e}_3$, la longitud no está bien definida; cualquiera es buena. Los mapas planos de la Tierra usan estas coordenadas, pero entonces los polos "son" las líneas horizontales extremas y las distancias no corresponden a las distancias reales en la esfera. Tampoco las líneas rectas en un mapa corresponden a las trayectorias más cortas en la esfera.

5.2.1 Líneas esféricas y polaridad



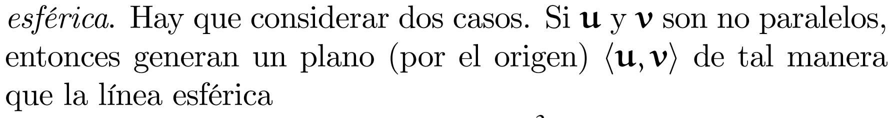
Consideremos dos puntos $\mathbf{u}$ y $\mathbf{v}$ en la esfera $\mathbb{S}^2$. ¿Cuál será la trayectoria más corta para ir de $\mathbf{u}$ a $\mathbf{v}$? Nos referimos a la trayectoria dentro de $\mathbb{S}^2$, es decir, sin salir nunca de la esfera, como si fuéramos una hormiga sobre ella o un ser humano en la Tierra. En $\mathbb{R}^3$ la trayectoria más corta es el segmento $\overline{\mathbf{uv}}$ entre los puntos; pero éste inmediatamente deja de estar en la esfera, y su interior cae dentro del interior de la bola sólida. Lo que más se parece al segmento $\overline{\mathbf{uv}}$ dentro de la esfera es su proyección desde el origen, su sombra en la esfera pensando que el centro de la esfera, el origen, es un foco. Pues entre más grande sea el radio de un círculo más se parece localmente a una recta y los círculos más grandes que hay en la esfera son los que tienen su centro en el origen, y que por tanto tienen radio uno.

No vamos a demostrar que las trayectorias más cortas (o eficientes) dentro de la esfera son los *círculos máximos* (*i.e.*, intersección de la esfera con planos por el origen) pues ello requiere herramientas de cálculo. Pero las vamos a tomar como definición de *líneas esféricas* (que, aunque suene contradictorio, viene de "líneas en la esfera") siguiendo la intuición de que son lo más parecido a las líneas rectas dentro de $\mathbb{S}^2$.

Definición 5.2.1 Una *línea* en la esfera, o una *línea esférica* es la intersección de $\mathbb{S}^2$ con un plano por el origen.

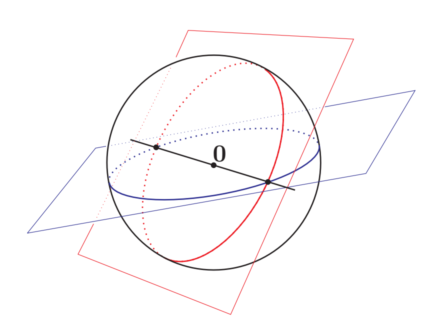
Nótese que cada línea esférica es (una copia de) el círculo unitario $\mathbb{S}^1$ pues consta de los vectores de norma 1 en un plano.

Veamos ahora que *por dos puntos* $\mathbf{u}$ *y* $\mathbf{v}$ *en la esfera* $\mathbb{S}^2$ *siempre pasa una línea esférica*. Hay que considerar dos casos. Si $\mathbf{u}$ y $\mathbf{v}$ son no paralelos, entonces generan un plano (por el origen) $\langle \mathbf{u}, \mathbf{v} \rangle$ de tal manera que la línea esférica



$$\xi = \langle \mathbf{u}, \mathbf{v} \rangle \cap \mathbb{S}^2$$

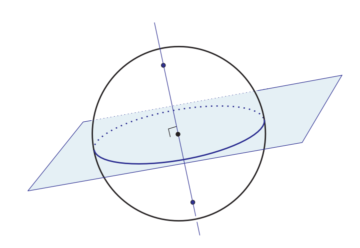
los contiene a ambos. Y si $\mathbf{u}$ y $\mathbf{v}$ son paralelos entonces están contenidos en una línea por el origen y (al ser de norma 1) o bien son iguales, o bien son *antípodas*, es decir, $\mathbf{u} = -\mathbf{v}$ ("forman un $\mathbb{S}^0$"); y en este caso, todas las líneas por uno también contienen el antípoda. Hemos demostrado que por cualquier par de puntos pasa una línea que además es única si y sólo si no son antípodas.



Esta condición es muy parecida a la de la geometría clásica del plano euclidiano donde por cualquier par de puntos pasa una recta. Pero a diferencia del plano euclidiano, en la esfera no hay paralelismo.

Cualquier par de líneas esféricas se intersectan en un pareja de puntos antípodas. Como los planos que las definen pasan por el origen, entonces se intersectan en una línea por él, que intersecta la esfera en una pareja de antípodas.

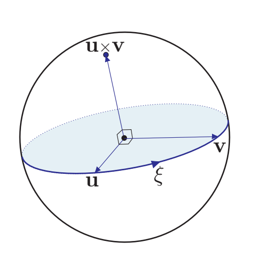
Polaridad



Puesto que en $\mathbb{R}^3$ tenemos que líneas y planos por el origen están naturalmente apareados por ortogonalidad, la esfera hereda este apareamiento que llamaremos *polaridad*:

Decimos que una línea esférica y un par antípoda en $\mathbb{S}^2$ son *polares* si el plano por el origen y la línea por el origen correspondientes son ortogonales. La polaridad da una correspondencia biyectiva natural entre los $\mathbb{S}^1$ y los $\mathbb{S}^0$ de $\mathbb{S}^2$. En la Tierra, por ejemplo, el ecuador es polar de los polos (norte y sur).

Si tomamos en cuenta la orientación de $\mathbb{R}^3$ podemos refinar nuestra noción de polaridad. A una línea esférica ξ le podemos dar dos orientaciones, es decir, maneras preferentes de viajar en ella. Si denotamos por $\vec{\xi}$ a la línea esférica ξ *orientada* (con una orientación dada), entonces de su par antípoda de polos podemos escoger uno al que llamaremos su *punto polar*: tomamos cualquier punto $\mathbf{u} \in \xi$, viajamos en la orientación preferida de $\vec{\xi}$ un cuarto de vuelta para llegar a un punto $\mathbf{v}$, y luego tomamos $\mathbf{u} \times \mathbf{v}$ (por la manera que escogimos $\mathbf{v}$ se tiene que $\mathbf{u} \times \mathbf{v} \in \mathbb{S}^2$) y el resultado no depende de $\mathbf{u}$. Geométricamente, y de acuerdo con nuestra convención de cómo tomar la base canónica, si con los dedos de la mano derecha emulamos, seguimos, la orientación de $\vec{\xi}$, el pulgar apunta hacia su punto polar; o bien, si caminamos sobre $\vec{\xi}$ de frente y de acuerdo con su orientación, a nuestra izquierda está su punto polar. Se tiene entonces que a las dos orientaciones de ξ corresponden sus dos puntos antípodas polares (pues $\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$); y a cada punto de $\mathbb{S}^2$ le corresponde una línea esférica orientada, llamada su *polar*.



5.2.2 Distancias y ángulos

Hemos usado intuitivamente que hay una distancia natural en la esfera $\mathbb{S}^2$, formalicémosla. Dados dos puntos $\mathbf{u}$ y $\mathbf{v}$ en $\mathbb{S}^2$, hay una línea esférica ξ que los contiene; como es un círculo de radio 1, el recorrido mínimo entre $\mathbf{u}$ y $\mathbf{v}$ a lo largo de ξ es precisamente el ángulo, medido en radianes, entre los vectores $\mathbf{u}$ y $\mathbf{v}$. Así que hay

que definir la *distancia en la esfera* de $\mathbf{u}$ a $\mathbf{v}$ como

$$\mathrm{d}_{\mathbb{S}^2}(\mathbf{u},\mathbf{v}) := \operatorname{ang}(\mathbf{u},\mathbf{v}).$$

Recordando la definición de ángulo entre vectores (1.10.2) que se deriva de la fórmula geométrica del producto interior, $\mathbf{u}\cdot\mathbf{v} = |\mathbf{u}|\,|\mathbf{v}|\cos(\operatorname{ang}(\mathbf{u},\mathbf{v}))$, y usando que $|\mathbf{u}| = |\mathbf{v}| = 1$ (pues están en $\mathbb{S}^2$) se obtiene que

$$\mathrm{d}_{\mathbb{S}^2}(\mathbf{u},\mathbf{v}) = \arccos(\mathbf{u}\cdot\mathbf{v}).$$

Cuando sea obvio que nos referimos a la distancia en $\mathbb{S}^2$, y no se confunda con la distancia en $\mathbb{R}^3$, olvidaremos el subíndice. En la práctica, la fórmula equivalente que más utilizaremos es

$$\mathbf{u}\cdot\mathbf{v} = \cos(\mathrm{d}_{\mathbb{S}^2}(\mathbf{u},\mathbf{v})).$$

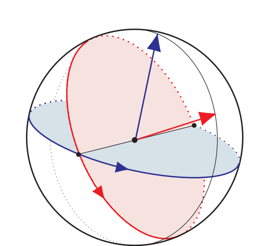
Como el arcocoseno toma valores en el intervalo $[0,\pi]$, se tiene entonces que para todo $\mathbf{u}$ y $\mathbf{v}$ en $\mathbb{S}^2$,

$$0 \le \mathrm{d}_{\mathbb{S}^2}(\mathbf{u},\mathbf{v}) \le \pi,$$

donde las igualdades, o extremos, se dan sólo cuando $\mathbf{u}$ y $\mathbf{v}$ son el mismo punto, y su distancia es 0, o bien cuando son antípodas y logran el máximo π. Y es claro también que la línea polar a un punto $\mathbf{u}$ consiste justo de los puntos a distancia $\pi/2$ de él (y que equidistan de su antípoda).

Para ver que esta distancia cumple con todas las de la ley (las condiciones del Teorema 1.11.1) nos falta probar la desigualdad del triángulo, $\mathrm{d}(\mathbf{u},\mathbf{v}) \le \mathrm{d}(\mathbf{u},\mathbf{w}) + \mathrm{d}(\mathbf{w},\mathbf{v})$. Pero conviene dejarlo para después, pues se usará la ley (esférica) de los cosenos en la que interviene la noción de ángulo.

Como sucedió en el plano, el ángulo entre dos líneas no está bien definido, pues se puede tomar cualquiera de los dos ángulos complementarios que definen. Pero esta ambigüedad se pierde si las líneas están orientadas y se mide entonces, e intuitivamente, el movimiento angular que gira una en la otra haciendo corresponder las orientaciones. Formalmente, consideremos dos rectas orientadas ξ y ζ (nótese que eliminamos la flechita de la notación, pues aquí estorba más de lo que ayuda) y sean, respectivamente, $\mathbf{u}$ y $\mathbf{v}$ sus puntos polares. Obsérvese que $\mathbf{u}$ y $\mathbf{v}$ están en la línea polar al par de antípodas $\xi\cap\zeta$, y que al mover $\mathbf{u}$ hacía $\mathbf{v}$ en esta línea, la polar a $\mathbf{u}$ gira justo alrededor de la intersección $\xi\cap\zeta$ para llevar a ξ, con la orientación adecuada, a ζ. De tal manera que debemos definir el *ángulo* entre ξ y ζ como



$$\operatorname{ang}(\xi,\zeta) := \mathrm{d}_{\mathbb{S}^2}(\mathbf{u},\mathbf{v}).$$

Para que esta fórmula quede únicamente en términos de las líneas orientadas, debemos introducir notación para la polaridad.

Notación 5.1 Si ξ es una línea esférica orientada, denotamos por $\xi^\perp$ (léase "ξ polar") a su punto polar; y para un punto $\mathbf{u} \in \mathbb{S}^2$ denotamos con $\mathbf{u}^\perp$ a su línea polar orientada. De tal manera que $\left(\xi^\perp\right)^\perp = \xi$ y $\left(\mathbf{u}^\perp\right)^\perp = \mathbf{u}$. (No hay que confundir la notación con el compadre ortogonal en $\mathbb{R}^2$, que ya no usaremos.)

Y entonces tenemos que si ξ y ζ son líneas orientadas, y $\mathbf{u}$ y $\mathbf{v}$ puntos en $\mathbb{S}^2$:

$$\operatorname{ang}(\xi, \zeta) = d_{\mathbb{S}^2}\left(\xi^\perp, \zeta^\perp\right) \quad \text{y} \quad d_{\mathbb{S}^2}(\mathbf{u}, \mathbf{v}) = \operatorname{ang}\left(\mathbf{u}^\perp, \mathbf{v}^\perp\right), \tag{5.5}$$

y además que

$$\mathbf{u} \in \xi \quad \Leftrightarrow \quad \xi^\perp \in \mathbf{u}^\perp.$$

Ángulos diédricos

Generalizando, se puede definir el ángulo entre dos planos para que el ángulo entre dos planos por el origen sea justo el ángulo entre las líneas esféricas correspondientes. De nuevo hay una ambigüedad, pues puede uno referirse a un ángulo o su complementario, que se resuelve orientando. Si un *plano orientado* es un plano junto con un lado positivo distinguido, *i.e.*, un vector normal, unitario digamos, especificado, entonces el ángulo entre dos planos orientados (*ángulo diédrico* o *diedro*, como a veces se le llama) es el ángulo entre sus vectores normales.

Otra manera de verlo, o medirlo, es tomando un plano Π perpendicular a los dos planos dados, es decir, ortogonal a su línea de intersección. En Π, los dos planos se intersectan en dos rectas; la orientación como lado preferido nos da líneas orientadas, y el ángulo diédrico entre los planos es el ángulo entre las líneas de intersección. O bien es el ángulo entre las líneas esféricas que definen al intersectarlos con una esfera centrada en su línea de intersección.

EJERCICIO 5.15 ¿Cómo son los círculos en la esfera? Es decir, ¿cómo son los conjuntos $\left\{\mathbf{x} \in \mathbb{S}^2 \mid d_{\mathbb{S}^2}(\mathbf{u}, \mathbf{x}) = r\right\}$ con $\mathbf{u} \in \mathbb{S}^2$ fijo (el centro del círculo) y $r > 0$ constante (el radio)?.

EJERCICIO 5.16 Demuestra que para cualquier par de puntos $\mathbf{u}, \mathbf{v} \in \mathbb{S}^2$ se tiene que

$$d_{\mathbb{S}^2}(\mathbf{u}, \mathbf{v}) + d_{\mathbb{S}^2}(\mathbf{u}, -\mathbf{v}) = \pi.$$

EJERCICIO 5.17 Sean ξ y ζ dos líneas esféricas. Demuestra que son ortogonales (*i.e.* que $\operatorname{ang}(\xi, \zeta) = \pi/2$ con cualquier orientación que se les dé a las líneas) si y sólo si $\xi^\perp \subset \zeta$, donde $\xi^\perp$ denota al par antípoda polar a ξ cuando ésta no tiene orientación preferida.

EJERCICIO 5.18 Demuestra que por cualquier punto $\mathbf{u} \in \mathbb{S}^2$ se puede trazar una línea ortogonal a una línea dada ξ; y que además ésta es única si y sólo si $\mathbf{u}$ no es polar de ξ.

EJERCICIO 5.19 Sean ξ_1 y ξ_2 dos líneas esféricas ortogonales. Demuestra que hay una única línea ξ_3 ortogonal a ambas. ¿Usaste el hecho de que fueran ortogonales?

EJERCICIO 5.20 Sean u y v dos puntos no antípodas en $\mathbb{S}^2$. Demuestra que la línea por u y v es $\left(u^\perp \cap v^\perp\right)^\perp$, dándole sentido a este abuso de notación.

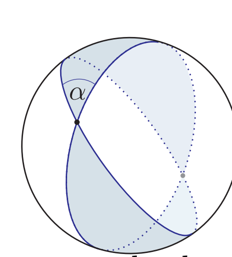
EJERCICIO 5.21 Demuestra que la distancia esférica se puede escribir en términos de la euclidiana como

$$d_{\mathbb{S}^2}(u, v) = 2 \text{ arcsen} \left(d_{\mathbb{R}^3}(u, v)/2\right)$$

y describe la función arcsen.

5.2.3 Área y triángulos

En la esfera, los ángulos también miden área. Si tomamos un ángulo entre dos rectas junto con su opuesto por el vértice, obtenemos dos gajos, llamémoslo un *bigajo*, determinado por el ángulo α entre las rectas. Al crecer α, el área del bigajo correspondiente, que denotaremos g_α, crece proporcionalmente —o mejor dicho, linealmente— pues el área del bigajo con ángulo $\alpha/2$, digamos, es justo la mitad del área de g_α ya que dos $g_{\alpha/2}$, sin traslaparse, cubren a g_α; y lo mismo sucede para cualquier entero n en vez de 2. Así que debe haber una constante c tal que



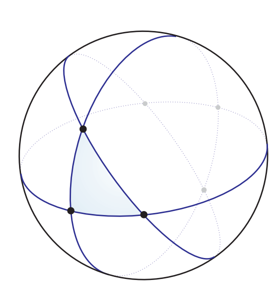
$$\mathcal{A}(g_\alpha) = c\,\alpha,$$

donde $\mathcal{A}$ representa el área. Puesto que el área del bigajo de ángulo π ya es el área de toda la esfera, también se tiene que

$$\mathcal{A}\left(\mathbb{S}^2\right) = c\,\pi.$$

Arquímedes calculó el valor de la constante c para cualquier esfera. Se tiene que $c = 4r^2$, donde r es el radio, y como en nuestro caso el radio es 1, la constante c que aparece arriba es 4. Aunque lo que nos interesa aquí es que exista, pues su cálculo preciso requiere de ideas de cálculo que están al margen.

Consideremos ahora una tercera línea que no pase por la intersección de las dos anteriores. No es difícil ver que entonces la esfera se parte en ocho *triángulos*, apareados en cuatro parejas de antípodas. Sea T uno de estos triángulos con α, β, γ sus ángulos internos, y sean $g_\alpha, g_\beta, g_\gamma$ los bigajos correspondientes. Si pintamos el bigajo g_α, luego todo el bigajo g_β y finalmente el g_γ, habremos pintado toda la esfera; pero tanto al triángulo T como a su antípoda le dimos tres manos, mientras que a los otros seis sólo una. Midiendo el área por la pintura que se uso (cubrir la esfera $\mathbb{S}^2$, dos manos adicionales para T y dos para su antípoda, que miden lo mismo) se obtiene que



$$\mathcal{A}(g_\alpha) + \mathcal{A}(g_\beta) + \mathcal{A}(g_\gamma) = \mathcal{A}\left(\mathbb{S}^2\right) + 4\,\mathcal{A}(T),$$

de donde, sustituyendo nuestras otras igualdades, concluimos que

$$\mathcal{A}(T) = \frac{c}{4}(\alpha + \beta + \gamma - \pi).$$

Como el área es estrictamente positiva, esto implica que la suma de los ángulos internos de un triángulo esférico es estrictamente mayor que π. Pero la fórmula anterior nos dice mucho más: que el exceso sobre π es proporcional al área del triángulo; si la suma se excede un poco es porque el triángulo es muy pequeño y se asemeja a uno plano. Para enunciar elegantemente el teorema que ya hemos demostrado, usaremos el valor de la constante que encontró Arquímedes.

Teorema 5.10 *La suma de los ángulos internos de un triángulo esférico es mayor que π; y el exceso sobre π es su área.* □

EJERCICIO 5.22 Demuestra que tres planos en $\mathbb{R}^3$ se intersectan en un único punto si y sólo si la suma de sus ángulos diédricos es mayor que π.

EJERCICIO 5.23 ¿Cuál es el máximo valor de la suma de los ángulos internos de un triángulo esférico? Da ejemplos de triángulos que se aproximen tanto como uno quiera a este valor.

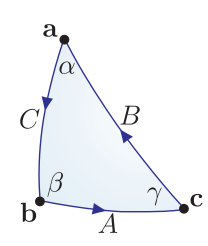
EJERCICIO 5.24 Demuestra que hay tres tipos de triángulos esféricos regulares (con sus tres ángulos iguales) con ángulo de la forma $2\pi/n$. En cada caso, ¿con cuántos de ellos se cubre la esfera?

EJERCICIO 5.25 Define polígono regular en la esfera y argumenta que para cada α con

$$\frac{n-2}{n}\pi < \alpha < \pi$$

hay un polígono regular (esférico) con n lados y ángulo interno α.

5.2.4 Trigonometría esférica



Sigamos analizando el triángulo T que tenemos entre manos. Si los ángulos internos determinan el área, también deben determinar la longitud de los lados. Denotemos con $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ los vértices del triángulo con ángulos correspondientes α, β, γ; y, abusando de la notación, denotemos con A, B, C tanto a las líneas orientadas de los lados opuestos como a las distancias correspondientes (*e.g.*, $A = d_{\mathbb{S}^2}(\mathbf{b}, \mathbf{c})$).

Lema 5.11 (Ley esférica de cosenos) *Con la notación anterior se tiene que*

$$\cos\alpha = \frac{\cos A - \cos B \cos C}{\operatorname{sen} B \operatorname{sen} C} \qquad y \qquad \cos A = \frac{\cos\alpha + \cos\beta\cos\gamma}{\operatorname{sen}\beta \operatorname{sen}\gamma}.$$

Demostración. La primera ecuación se sigue de las dos interpretaciones que le podemos dar a $(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{a} \times \mathbf{c})$. Primero, con un desarrollo similar al final de la

demostración del Teorema 5.2, se obtiene

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{a}\times\mathbf{c}) &= \mathbf{a}\cdot(\mathbf{b}\times(\mathbf{a}\times\mathbf{c}))\\ &= -\,\mathbf{a}\cdot((\mathbf{a}\times\mathbf{c})\times\mathbf{b})\\ &= -\,\mathbf{a}\cdot((\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}-(\mathbf{c}\cdot\mathbf{b})\,\mathbf{a})\\ &= (\mathbf{c}\cdot\mathbf{b})\,(\mathbf{a}\cdot\mathbf{a})-(\mathbf{a}\cdot\mathbf{b})\,(\mathbf{a}\cdot\mathbf{c})\\ &= \cos A-\cos C\,\cos B. \end{aligned} \tag{5.6}$$

Por otro lado, usando el Teorema 5.2 se tiene que $|\mathbf{a}\times\mathbf{b}| = |\mathbf{a}|\,|\mathbf{b}|\operatorname{sen} C = \operatorname{sen} C$; pero además sabemos que $\mathbf{a}\times\mathbf{b}$ está en la dirección de $C^{\perp}$ (aquí estamos interpretando por primera vez a C como línea orientada y tomando su punto polar). Entonces tenemos que

$$\mathbf{a}\times\mathbf{b} = (\operatorname{sen} C)\;C^{\perp}.$$

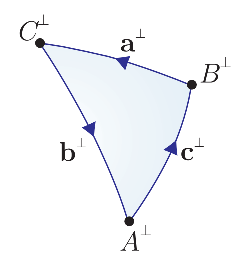
Análogamente, cuidando la orientación, tenemos que $\mathbf{c}\times\mathbf{a} = (\operatorname{sen} B)\;B^{\perp}$, y por tanto,

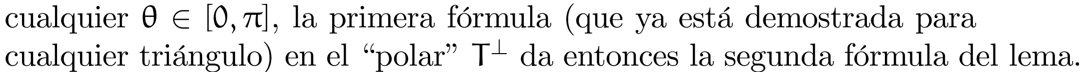
$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{a}\times\mathbf{c}) &= -\,(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{a})\\ &= -\left((\operatorname{sen} C)\;C^{\perp}\right)\cdot\left((\operatorname{sen} B)\;B^{\perp}\right)\\ &= (\operatorname{sen} C\operatorname{sen} B)\left(-C^{\perp}\cdot B^{\perp}\right)\\ &= \operatorname{sen} C\operatorname{sen} B\cos\alpha, \end{aligned} \tag{5.7}$$

donde estamos usando que para medir el ángulo interior, hay que cambiarle la orientación a alguno de los lados.

De (5.6) y (5.7) se sigue la primera fórmula del lema.

La segunda se sigue de la primera por polaridad. Si consideramos los tres puntos polares a los lados orientados del triángulo T, obtenemos un triángulo cuyos lados orientados son las líneas polares a los tres vértices de T. Y en este triángulo, "polar" podríamos llamarlo y que denotamos $T^{\perp}$, lo que miden los lados es lo que miden los ángulos externos de T por (5.5). Si denotamos $\bar{\alpha}$ al *complemento* de $\alpha\in[0,\pi]$, es decir, $\bar{\alpha} := \pi-\alpha$, entonces los lados de $T^{\perp}$ miden $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$ correspondiendo a las líneas $\mathbf{a}^{\perp}$, $\mathbf{b}^{\perp}$, $\mathbf{c}^{\perp}$ respectivamente. Análogamente, los ángulos internos de $T^{\perp}$ miden $\bar{A}$, $\bar{B}$, $\bar{C}$, respectivamente. Puesto que $\cos\bar{\theta} = -\cos\theta$ y $\operatorname{sen}\bar{\theta} = \operatorname{sen}\theta$ para cualquier $\theta\in[0,\pi]$, la primera fórmula (que ya está demostrada para cualquier triángulo) en el "polar" $T^{\perp}$ da entonces la segunda fórmula del lema. □

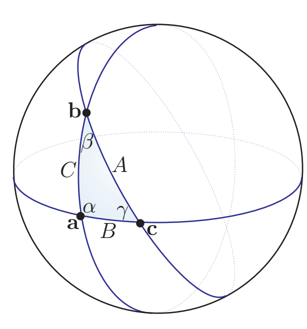




Como corolario (de la segunda fórmula) obtenemos que los ángulos de un triángulo esférico determinan la longitud de sus lados, a diferencia del plano euclidiano, y entonces no hay "semejanza" en los triángulos esféricos. Pero, como antes, los lados de un triángulo determinan a sus ángulos (por la primera fórmula).

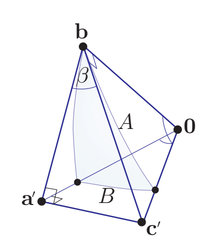
Lema 5.12 (Ley esférica de senos) *Con la notación del lema anterior se tiene*

$$\frac{\operatorname{sen} A}{\operatorname{sen}\alpha} = \frac{\operatorname{sen} B}{\operatorname{sen}\beta} = \frac{\operatorname{sen} C}{\operatorname{sen}\gamma}.$$



Demostración. Aunque hay una demostración algebraica de la ley de los senos usando la de los cosenos, preferimos una geométrica que sigue en parte la versión euclidiana.

Primero hay que ver el caso en que el triángulo es rectángulo; digamos que $\alpha = \pi/2$; y para fijar ideas, supongamos que el punto $\mathbf{a}$ coincide con el primer vector canónico, que el punto $\mathbf{b}$ está en el plano xz y que el punto $\mathbf{c}$ está en el plano xy. Consideremos el



tetraedro (dícese de un cuerpo acotado por cuatro planos en $\mathbb{R}^3$) definido por los tres planos de las líneas (esféricas) del triángulo T (que se intersectan en el origen), y el plano Π tangente a la esfera en el punto $\mathbf{b}$, dado por la ecuación

$$\mathbf{b} \cdot \mathbf{x} = 1.$$

Este tetraedro tiene cuatro triángulos como caras. Nótese que las longitudes de los tres lados de nuestro triángulo esférico original T son ahora los ángulos de los tres triángulos que tienen como vértice el origen, y que el ángulo β es el ángulo en el vértice $\mathbf{b}$ del triángulo en el plano Π (justo porque Π es tangente a la esfera en $\mathbf{b}$). Los otros dos vértices del tetraedro son $\mathbf{a}' = (\mathbf{a} \cdot \mathbf{b})^{-1}\, \mathbf{a}$ y $\mathbf{c}' = (\mathbf{c} \cdot \mathbf{b})^{-1}\, \mathbf{c}$. No es difícil ver que los cuatro triángulos, caras, del tetraedro son rectángulos; dos en el vértice $\mathbf{b}$ (por ser perpendicular $\mathbf{b}$ a Π) y dos en el vértice $\mathbf{a}'$ (aquí es donde hay que usar que $\alpha = \pi/2$). Entonces tenemos las siguientes igualdades (con distancias euclidianas):

$$\operatorname{sen}\beta = \frac{d(\mathbf{a}', \mathbf{c}')}{d(\mathbf{b}, \mathbf{c}')}\ , \quad \operatorname{sen} B = \frac{d(\mathbf{a}', \mathbf{c}')}{d(\mathbf{0}, \mathbf{c}')}\ , \quad \operatorname{sen} A = \frac{d(\mathbf{b}, \mathbf{c}')}{d(\mathbf{0}, \mathbf{c}')}.$$

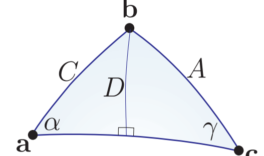
De donde se concluye que

$$\operatorname{sen}\beta = \frac{\operatorname{sen} B}{\operatorname{sen} A}.$$

Nuestro argumento funciona para valores pequeños de A, B y C (como en el dibujo), pero no es difícil ver que también funciona para $B \geq \pi/2$ y $C \geq \pi/2$.

Para el caso general, se traza la perpendicular de un vértice, $\mathbf{b}$ digamos, al lado opuesto; y si D es la altura entonces se obtiene, por el caso anterior en los dos triángulos rectángulos que se forman, que $\operatorname{sen}\alpha \operatorname{sen} C = \operatorname{sen} D = \operatorname{sen}\gamma \operatorname{sen} A$, de donde

$$\frac{\operatorname{sen} A}{\operatorname{sen}\alpha} = \frac{\operatorname{sen} C}{\operatorname{sen}\gamma},$$



y el lema se sigue por simetría. □

EJERCICIO 5.26 Demuestra que en un triángulo (esférico, se sobreentiende) sus tres ángulos son iguales si y sólo si sus tres lados son iguales. Tales triángulos se llaman *regulares* o *equiláteros*. ¿Cuántos de ellos hay? ¿Puedes parametrizarlos? ¿Por su ángulo? ¿Por su lado? ¿Por su radio? ¿Puedes dar las fórmulas que relacionan estos parámetros?

EJERCICIO 5.27 Demuestra que si un triángulo tiene dos ángulos iguales entonces sus lados opuestos son iguales. E inversamente, si tiene dos lados iguales, los ángulos opuestos son iguales. A estos triángulos se les llama *isósceles*.

EJERCICIO 5.28 Demuestra que la línea que va del vértice distinto de un triángulo isósceles al punto medio de su lado opuesto es perpendicular a éste. Al segmento en cuestión se le llama la *altura* del triángulo.

EJERCICIO 5.29 Demuestra que el lugar geométrico de los puntos (en $\mathbb{S}^2$, se sobreentiende) que equidistan de dos puntos dados es la línea perpendicular al segmento que los une y que pasa por su punto medio. Se le llama la *mediatriz*.

EJERCICIO 5.30 ¿Cuánto miden el área, el lado y el diámetro de un cuadrado regular con ángulo interno $2\pi/3$?

EJERCICIO 5.31 ¿Puedes dar una cota superior para la suma de los lados de un triángulo? La inferior es 0.

5.3 Isometrías de la esfera ($\mathbf{O}(3)$)

Estudiaremos ahora las transformaciones rígidas de la esfera. Como sabemos, una función $f : \mathbb{S}^2 \to \mathbb{S}^2$ que preserva distancias se llama *isometría*. Veremos primero que las isometrías de la esfera corresponden justo a las transformaciones ortogonales de $\mathbb{R}^3$ que habíamos denotado $\mathbf{O}(3)$.

Una transformación ortogonal de $\mathbb{R}^3$, por definición, preserva producto interior. Por tanto preserva norma y manda a la esfera unitaria en sí misma, pues es el conjunto de vectores unitarios. Además, como la distancia en $\mathbb{S}^2$ se definió con el producto interior (es "distancia angular"), también preserva distancias, por lo que su restricción a $\mathbb{S}^2$ es una isometría.

Dibujo...

Inversamente, si $f : \mathbb{S}^2 \to \mathbb{S}^2$ es una isometría queremos demostrar que es la restricción de una transformación ortogonal. Sean $\mathbf{u} = f(\mathbf{e}_1), \mathbf{v} = f(\mathbf{e}_2), \mathbf{w} = f(\mathbf{e}_3)$. Como el *triángulo canónico* $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ tiene sus tres lados de tamaño $\pi/2$ (igual que sus tres ángulos) y f preserva distancias, entonces el triángulo $\mathbf{u}, \mathbf{v}, \mathbf{w}$ tiene las mismas características, que equivalen a decir que $\mathbf{u}, \mathbf{v}, \mathbf{w}$ forman una base ortonormal de $\mathbb{R}^3$. Nos falta entonces demostrar que para cualquier $\mathbf{x} = (x, y, z) \in \mathbb{S}^2$ se tiene que

$$f(\mathbf{x}) = x\,\mathbf{u} + y\,\mathbf{v} + z\,\mathbf{w},$$

es decir, que f es lineal, pues entonces f será la restricción de la transformación ortogonal asociada a la matriz $(\mathbf{u}, \mathbf{v}, \mathbf{w}) \in \mathbf{O}(3)$.

La demostración de que f es lineal (análoga a la que dimos en el Capítulo 3 para $\mathbb{R}^2$) consiste en reinterpretar geométricamente el teorema de bases ortonormales en $\mathbb{R}^3$ que en la Sección 1.10 dejamos como ejercicio. Por su importancia, lo volvemos a enunciar y demostrar.

Teorema 5.13 (Bases ortonormales) *Sea $\mathbf{u}, \mathbf{v}, \mathbf{w}$ una base ortonormal de $\mathbb{R}^3$, entonces para cualquier $\mathbf{x} \in \mathbb{R}^3$ se tiene que*

$$\mathbf{x} = (\mathbf{x} \cdot \mathbf{u})\ \mathbf{u} + (\mathbf{x} \cdot \mathbf{v})\ \mathbf{v} + (\mathbf{x} \cdot \mathbf{w})\ \mathbf{w}.$$

Demostración. Para resolver el sistema de ecuaciones

$$\mathbf{x} = s\,\mathbf{u} + t\,\mathbf{v} + r\,\mathbf{w}$$

con incógnitas s, t, r, tomamos el producto punto con los vectores dándonos los valores de las tres variables por ser $\mathbf{u}, \mathbf{v}, \mathbf{w}$ una base ortonormal. Así, por ejemplo, tomando el producto punto con $\mathbf{u}$ se tiene

$$\mathbf{x} \cdot \mathbf{u} = s\ (\mathbf{u} \cdot \mathbf{u}) + t\ (\mathbf{v} \cdot \mathbf{u}) + r\ (\mathbf{w} \cdot \mathbf{u}) = s$$

y análogamente $t = \mathbf{x} \cdot \mathbf{v}$ y $r = \mathbf{x} \cdot \mathbf{w}$. □

Regresando a nuestra isometría f y un punto cualquiera $\mathbf{x} = (x, y, z) \in \mathbb{S}^2$, se tiene que $x = \mathbf{x} \cdot \mathbf{e}_1$ mide esencialmente la distancia de $\mathbf{x}$ al vector canónico $\mathbf{e}_1$ (es su coseno) y esta distancia tiene que ser la misma que de $f(\mathbf{x})$ a $\mathbf{u} = f(\mathbf{e}_1)$. Más formalmente,

$$f(\mathbf{x}) \cdot \mathbf{u} = \cos\left(d_{\mathbb{S}^2}\ (f(\mathbf{x}), f(\mathbf{e}_1))\right) = \cos\left(d_{\mathbb{S}^2}\ (\mathbf{x}, \mathbf{e}_1)\right) = \mathbf{x} \cdot \mathbf{e}_1.$$

Análogamente $f(\mathbf{x}) \cdot \mathbf{v} = \mathbf{x} \cdot \mathbf{e}_2$ y $f(\mathbf{x}) \cdot \mathbf{w} = \mathbf{x} \cdot \mathbf{e}_3$, de donde, por el Teorema de bases ortonormales, obtenemos que

$$\begin{aligned} f(\mathbf{x}) &= (f(\mathbf{x}) \cdot \mathbf{u})\ \mathbf{u} + (f(\mathbf{x}) \cdot \mathbf{v})\ \mathbf{v} + (f(\mathbf{x}) \cdot \mathbf{w})\ \mathbf{w} \\ &= (\mathbf{x} \cdot \mathbf{e}_1)\ \mathbf{u} + (\mathbf{x} \cdot \mathbf{e}_2)\ \mathbf{v} + (\mathbf{x} \cdot \mathbf{e}_3)\ \mathbf{w} \\ &= x\,\mathbf{u} + y\,\mathbf{v} + z\,\mathbf{w}, \end{aligned}$$

y queda demostrado que toda isometría de $\mathbb{S}^2$ es la restriccion de una transformación ortogonal de $\mathbb{R}^3$.

Nótese que en el fondo esta demostración (y el teorema) dice que todo punto de la esfera queda determinado por sus distancias a los tres vectores canónicos y bajo una isometría tiene que ir al punto con las mismas distancias a la imagen del triángulo canónico. Dibujo...

Podemos formalizar lo que intuitivamente es claro, que la esfera es homogénea (como espacio) y super-simétrica. Un punto puede ir a cualquier otro punto; si

además especificamos líneas orientadas por ellos, las podemos hacer coincidir, y nos queda la libertad de escoger cómo hacer corresponder a los dos lados de las líneas. Pues podemos pensar que una base ortonormal consiste en escoger un punto $\mathbf{u}$, una línea orientada que pasa por él (el segundo elemento de la base está a distancia $\pi/2$ en esa dirección) y luego un lado de la línea (el positivo nos da al producto cruz como tercer elemento de la base y el negativo a su antípoda).

También es claro que ahora tenemos la noción correcta de *igualdad* de triángulos: dos lo son si hay una isometría que mande uno en el otro. Y teoremas tales como que dos triángulos son iguales si y sólo si sus lados o sus ángulos (o un lado y sus dos ángulos adyacentes, o un ángulo y sus dos lados adyacentes) son iguales. Las demostraciones consisten en ir moviendo la esfera para que datos correspondientes vayan correspondiendo conjuntista, o estrictamente.

EJERCICIO 5.32 Demuestra que las clases de isometría de triángulos esféricos corresponden a las ternas $\alpha \leq \beta \leq \gamma$ con $0 < \alpha, \beta, \gamma < \pi$, tales que $\alpha + \beta + \gamma < 2\pi$, donde α, β, γ son los lados.

EJERCICIO 5.33 Decimos que dos triángulos esféricos están en la misma *familia* si ambos se pueden dibujar con tres líneas (recuerda la Sección 5.2.3 de áreas). ¿Cómo se relacionan los lados de triángulos en una familia? Demuestra que las familias tienen 4, 2 o un sólo elemento, donde la de un sólo elemento es la de los triángulos de bases ortonormales.

5.3.1 Rotaciones (SO(3))

A las matrices ortogonales con determinante 1 se les denota $\mathbf{SO}(3)$ por sus siglas en ingles *special orthogonal group.* Corresponden a las isometrías de $\mathbb{S}^2$ que preservan orientación, y que mandan al lado positivo de una línea orientada en el lado positivo de su imagen, es decir, que preservan el punto polar de las líneas orientadas. Las que invierten orientación (con determinante -1) no forman un grupo y mandan al punto polar de una línea orientada en el antípoda del polar de su imagen.

Ejemplos de isometrías que preservan orientación son las *rotaciones*, que fijan un punto (y su antípoda) y giran la esfera alrededor de ese eje. El teorema principal de esta sección es que todas son así; coloquialmente es el "Teorema del gol": después del primer gol, el balón regresa al centro de la cancha e independientemente de lo que le haya sucedido en el ínterin, un punto regresa justo a su lugar y el balón solamente giró alrededor de él.

Antes de demostrar el teorema, observemos que las "traslaciones" en la esfera también son rotaciones. Una traslación es "empujar" a lo largo de un línea; en la esfera una línea es un círculo y "empujar" en él es rotarlo, pero sus puntos polares se mantienen fijos, equivale a rotar en ellos. Podemos especificar que rotar en un punto $\mathbf{u}$ un ángulo α rota a su polar, como línea orientada, un ángulo α; denotemos esta

isometría $\rho_{\alpha,\mathbf{u}}$. Y entonces tenemos que

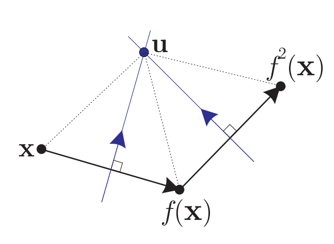
$$\rho_{\alpha,\mathbf{u}} = \rho_{-\alpha,-\mathbf{u}},$$

pues $-\mathbf{u}$ le da la orientación contraria a la línea polar. Por ejemplo, e identificando una transformación con su matriz asociada, se tiene

$$\rho_{\pi/2,\mathbf{e}_3} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Teorema 5.14 *Toda isometría de $\mathbb{S}^2$ que preserva orientación es una rotación.*

Demostración. Sea $f : \mathbb{S}^2 \to \mathbb{S}^2$ una isometría que preserva orientación. Tenemos que encontrar un punto fijo de f, es decir, un punto $\mathbf{u}$ tal que $f(\mathbf{u}) = \mathbf{u}$, y con eso habremos terminado; pues entonces manda su línea polar orientada en sí misma y el ángulo de la rotación queda determinado ahí.

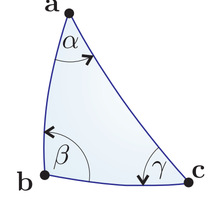


Sea $\mathbf{x}$ cualquier punto de $\mathbb{S}^2$. Podemos suponer que $f(\mathbf{x}) \neq \mathbf{x}$ pues si no ya acabamos. Consideremos el segmento (orientado) de $\mathbf{x}$ a $f(\mathbf{x})$ y volvamos a aplicarle la función f: va a dar a un segmento de $f(\mathbf{x})$ a $f^2(\mathbf{x})$ que mide lo mismo. Consideremos ahora las mediatrices de estos dos segmentos en la dirección positiva y sea $\mathbf{u}$ su primer punto de intersección. Como f preserva orientación, por construcción se tiene que $f(\mathbf{u}) = \mathbf{u}$ (el triángulo isósceles $\mathbf{x}, f(\mathbf{x}), \mathbf{u}$ tiene que ir al $f(\mathbf{x}), f^2(\mathbf{x}), \mathbf{u}$).

Nótese que la construcción es básicamente igual a la euclidiana, donde el caso en que las mediatrices no se intersectan corresponde a la translación. En la esfera como siempre se intersectan no hay esta excepción. Pero hicimos una trampa (no esencial), o bien nos falta analizar el caso en que las dos mediatrices son la misma recta (y el punto $\mathbf{u}$ no queda bien definido). Hay dos posibilidades: que el punto medio del segmento regrese a su lugar, y éste es el punto fijo de una rotación de π; o bien que los dos puntos medios sean distintos y generen a la mediatriz común, en cuyo caso el punto medio del segmento que los une es el punto medio de una rotación de π. □

La composición de dos rotaciones vuelve a ser una rotación, en dónde y de a cuánto nos lo dice el siguiente lema cuya idea es la misma del final de la Sección 3.8.1 sobre rotaciones euclidianas. La diferencia es que ahora los ángulos no se suman, pero sí se pueden obtener geométricamente.

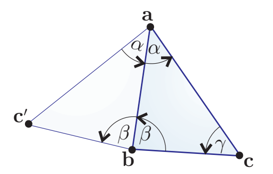
Lema 5.15 *Si tenemos un triángulo $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ con ángulos internos orientados correspondientes α, β, γ, entonces*



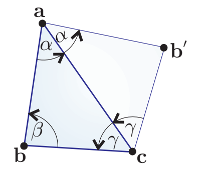
$$\rho_{2\alpha,\mathbf{a}} \circ \rho_{2\beta,\mathbf{b}} = \rho_{-2\gamma,\mathbf{c}}.$$

checar signo comp de ec.

Demostración.



Sea $f = \rho_{2\alpha,a} \circ \rho_{2\beta,b}$. Consideremos el reflejado del triángulo en el lado ab y sea c' su nuevo vértice. Como $d(b,c) = d(b,c')$ se tiene por construcción que $\rho_{2\beta,b}(c) = c'$, y análogamente que $\rho_{2\alpha,a}(c') = c$; entonces $c = f(c)$ es el centro de la rotación f.



Consideremos ahora el reflejado del triángulo en el lado ac y sea b' su nuevo vértice. Como $\rho_{2\beta,b}$ fija a b y $d(a,b) = d(a,b')$, se tiene que $f(b) = \rho_{2\alpha,a} \circ \rho_{2\beta,b}(b) = \rho_{2\alpha,a}(b) = b'$. Entonces el ángulo de la rotación f es el ángulo dirigido $\angle bcb' = -2\gamma$. Lo cual completa la demostración del lema. □

EJERCICIO 5.34 Demuestra que si $d_{\mathbb{S}^2}(u,v) = \pi/2$ entonces

$$\rho_{\pi,v} \circ \rho_{\pi,u} = \rho_{\pi,u\times v}.$$

EJERCICIO 5.35 ¿Cuál es la matriz asociada a la rotación de $2\pi/3$ en el punto $1/\sqrt{3}(1,1,1)$? Demuestra que es $\rho_{\pi/2,e_2} \circ \rho_{\pi/2,e_1}$. Observa que en las rotaciones de la esfera los ángulos de rotación no se suman al componer.

EJERCICIO 5.36 Demuestra que si una rotación manda a algún punto u en su antípoda entonces es una rotación de ángulo π con centro en un punto de $u^\perp$. Y más en general:

EJERCICIO 5.37 Demuestra que si f es una rotación de ángulo α entonces

$$\alpha = \max\left\{d_{\mathbb{S}^2}(x, f(x)) \mid x \in \mathbb{S}^2\right\}.$$

5.3.2 Pasos

Como corolario al teorema del gol, obtendremos que las isometrías de la esfera que invierten orientación siempre son pasos (reflejar en una línea y luego trasladar a lo largo de ella). El ingrediente nuevo que usaremos es la existencia de una isometría distinguida que invierte orientación. Es la función antípoda, que denotaremos con

$$a : \mathbb{S}^2 \to \mathbb{S}^2$$
$$a(x) = -x.$$

Puesto que su matriz asociada es $-I$, efectivamente invierte orientación y además conmuta con cualquier isometría pues éstas son lineales y la antípoda consiste en multiplicar por -1.

Teorema 5.16 *Toda isometría de $\mathbb{S}^2$ que invierte orientación es un paso.*

...Dibujo

Demostración. Sea $f : \mathbb{S}^2 \to \mathbb{S}^2$ una isometría que invierte orientación. Tenemos que $a \circ f = f \circ a$ preserva orientación y por el Teorema del Gol (5.14) es una rotación en algún punto, $\mathbf{u}$ digamos. Como $a \circ a = \mathrm{id}_{\mathbb{S}^2}$ entonces $f = a \circ (a \circ f)$ cumple que $f(\mathbf{u}) = -\mathbf{u}$ (pues $(a \circ f)(\mathbf{u}) = \mathbf{u}$); es decir, f permuta los puntos antípodas $\mathbf{u}$ y $-\mathbf{u}$. Entonces f deja invariante a la línea polar a $\mathbf{u}$, es decir $f\left(\mathbf{u}^\perp\right) = \mathbf{u}^\perp$ (pues está formada de los puntos a distancia $\pi/2$ de $\mathbf{u}$ y $-\mathbf{u}$); pero además f preserva la orientación de $\mathbf{u}^\perp$ pues si la invirtiera, como ya manda a $\mathbf{u}$ en $-\mathbf{u}$, preservaría la orientación de la esfera. Entonces f restringida a la línea $\mathbf{u}^\perp$ es una rotación (una traslación del círculo). Por lo tanto, f es un paso a lo largo de la línea $\mathbf{u}^\perp$. □

Aunque ya quedó demostrado el teorema, vale la pena recapitular en su demostración. Si la rotación $a \circ f$ en el punto $\mathbf{u}$ es de un ángulo α, entonces f, que se obtiene aplicándole la antipoda a esta rotación, vuelve a rotar un ángulo π a la línea $\mathbf{u}^\perp$, pues la función antípoda en un círculo cualquiera consiste en rotarlo π. Entonces el paso f es de tamaño $\pi + \alpha$; cuando $\alpha = \pi$ el paso es de 0 $(= 2\pi)$, o sea f es una reflexión, y cuando $\alpha = 0$ $(a \circ f = \mathrm{id}_{\mathbb{S}^2})$ entonces el paso es de π y $f = a$. Así como la identidad es la rotación de 0 alrededor de cualquier punto, resulta que la función antípoda es el paso de π en cualquier línea.

EJERCICIO 5.38 Demuestra que la composición de dos reflexiones en líneas con un ángulo α es una rotación en la intersección de un ángulo 2α. Concluye que toda isometría puede escribirse como una composición de tres reflexiones a lo más.

EJERCICIO 5.39 Demuestra que la composición de tres reflexiones en líneas mutuamente ortogonales es la función antípoda.

EJERCICIO 5.40 Demuestra que para cualquier $A \in \mathbf{O}(3)$ existen $B \in \mathbf{O}(3)$ y $\alpha \in [0, 2\pi]$ tales que

$$BAB^\top = \begin{pmatrix} \cos\alpha & -\operatorname{sen}\alpha & 0 \\ \operatorname{sen}\alpha & \cos\alpha & 0 \\ 0 & 0 & \pm 1 \end{pmatrix}.$$

5.4 *Simetría esférica

En esta sección estudiamos las posibles simetrías de los cuerpos en el espacio. Como en el plano, a cualquier cuerpo (o figura tridimensional) $F \subset \mathbb{R}^3$ le podemos asociar un grupo de isometrías que consiste de todas sus simetrías

$$\mathrm{Sim}(F) = \left\{ f \in \mathbf{Iso}(\mathbb{R}^3) \mid f(F) = F \right\}.$$

Por ejemplo, un trompo T tiene simetría *rotacional*, lo que quiere decir que $\mathrm{Sim}(T)$ es una copia de $\mathbf{O}(2)$ (y estrictamente $\mathbf{O}(2)$ si ponemos a su eje de simetría en el eje z);

Dibujos d un trompo un diábol girando

tiene las simetrías de un círculo y cualquiera de sus tajadas perpendiculares al eje es un círculo centrado en él. Eso es lo que hace interesantes sus propiedades mecánicas: el trompo es (idealmente) indistinguible de cualquiera de sus rotaciones. Su grupo de simetrías puede crecer si tuviera un plano de simetría donde lo de abajo es "igual" a lo de arriba (un diábolo o un platillo volador, por ejemplo); entonces todos los pasos (la reflexión seguida de rotaciones) en este plano se vuelven automáticamente simetrías y además aparecen rotaciones de π en ejes de este plano (estamos hablando de las matrices que aparecen en el último ejercicio y las que tienen una reflexión en la esquina). No es difícil intuir que si el cuerpo tiene alguna simetría extra (que necesariamente mueve a su eje, porque las anteriores son todas las que lo dejan en su lugar), entonces su grupo de simetrías crece de golpe hasta generar todo $\mathbf{O}(3)$ y estamos hablando entonces de una bola, un balín, una pelota, o bien la esfera $\mathbb{S}^2$. Estos son los posibles grupos de simetrías continuos; en el resto de la sección estudiaremos los finitos.

Nótese que si G es un subgrupo finito de isometrías de $\mathbb{R}^3$, entonces la misma demostración que dimos para el plano en el Paso 1 del Teorema de Leonardo (Sección 3.9, Teorema 3.26) prueba que tiene un centro de simetría (el centro de masas del cuerpo) que podemos asumir que es el origen. Entonces es equivalente pensar que G está actuando en la esfera $\mathbb{S}^2$ y hacer ahí nuestros razonamientos; podemos pensar bidimensionalmente y que estamos hablando de las simetrías de alguna figura F dibujada en la esfera. Así lo haremos.

Sea $\mathsf{G} \subset \mathbf{O}(3)$ un subgrupo finito. Queremos ver quién puede ser G. Como en el plano, G tiene asociados ciertos puntos distinguidos o *singulares* y con base en ellos detectaremos las posibilidades de G. Dado un punto cualquiera $x \in \mathbb{S}^2$ tenemos un subgrupo de G llamado su *estabilizador*

$$\mathbf{St}_{\mathsf{G}}(x) := \{g \in \mathsf{G} \mid g(x) = x\},$$

formado por los elementos de G que dejan fijo a x; y diremos que x es *singular* (para G) si su estabilizador es no trivial. Cualquier punto tiene la identidad en su estabilizador, pero si tiene algo más entonces es singular.

Dibujos...

Para cualquier punto singular (de G), $\nu \in \mathbb{S}^2$, su estabilizador es cíclico o diédrico. Pues si $g \in \mathbf{St}_{\mathsf{G}}(\nu)$ entonces g (pensado como transformación de $\mathbb{R}^3$) manda al plano ortogonal a ν (y al tangente a la esfera por ν) en sí mismo, y entonces podemos pensar que $\mathbf{St}_{\mathsf{G}}(\nu)$ es un grupo finito de isometrías del plano y por el Teorema de Leonardo es cíclico o diédrico. Si $\mathbf{St}_{\mathsf{G}}(\nu)$ es cíclico de orden n, diremos que ν es un *centro de simetría rotacional de orden* n; y si es el diédrico de orden n, que es un *centro de simetría caleidoscópica de orden* n. En el primer caso, $\mathbf{St}_{\mathsf{G}}(\nu)$ está generado por la rotación en ν de ángulo $2\pi/n$, y en el segundo $\mathbf{St}_{\mathsf{G}}(\nu)$ está generado por dos reflexiones cuyas líneas espejo se intersectan en ν con ángulo π/n.

A cualquier punto $x \in \mathbb{S}^2$ también podemos asociarle un subconjunto de $\mathbb{S}^2$ llamado su *órbita*

$$\mathsf{G}x := \{f(x) \mid f \in \mathsf{G}\},$$

que consta de todas las imágenes de x bajo los elementos de G. Demostraremos a continuación que *la órbita de un punto consta de puntos del mismo tipo*, es decir, que si x es un centro de simetría rotacional (o caleidoscópica) de orden n, entonces los puntos de su órbita también lo son. El truco es la vieja conocida conjugación.

Si $f \in G$ es tal que $f(x) = y$ (si y está en la órbita de x), entonces la función "conjugar por f" definida por

$$\kappa_f : G \to G$$
$$\kappa_f(g) = f^{-1} \circ g \circ f$$

manda al estabilizador de y en el de x. Pues si $g \in \mathbf{St}_G(y)$, por definición $g(y) = y$, y por tanto

$$\left(f^{-1} \circ g \circ f\right)(x) = \left(f^{-1} \circ g\right)(f(x)) = f^{-1}(g(y)) = f^{-1}(y) = x,$$

lo cual implica que $\kappa_f(g) \in \mathbf{St}_G(x)$. Es fácil ver que conjugar por f es una biyección (su inverso es conjugar por f^{-1}) y que preserva el producto del grupo ($\kappa_f(g \circ g') = \kappa_f(g) \circ \kappa_f(g')$), así que $\mathbf{St}_G(y)$ y $\mathbf{St}_G(x)$ son *isomorfos* (el mismo tipo de grupo).

EJERCICIO 5.41 Con la notación de arriba, demuestra que

$$\sharp(G) = \sharp(Gx)\,\sharp(\mathbf{St}_G(x)).$$

Donde $\sharp$ significa "*cardinalidad*", es decir, número de elementos de un conjunto.

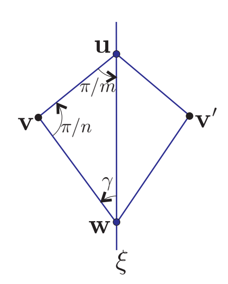
5.4.1 Subgrupos finitos de SO(3)

Consideremos primero el caso en que el subgrupo finito de isometrías de la esfera, G, preserva orientación; es decir, que todos sus elementos lo hacen, o, dicho de otra manera, que $G \subset \mathbf{SO}(3)$. Supongamos que G no es trivial, entonces tiene algún elemento f que no es la identidad. Por el Teorema 5.14, f es una rotación en un punto v, y entonces v es un punto singular para G, pues $f(v) = v$ y $f \neq \mathrm{id}_{\mathbb{S}^2}$. El estudio de estos puntos singulares nos llevará a determinar G.

Ya vimos que G tiene algún punto singular v. Éste es un centro de simetría rotacional de orden n, digamos, y entonces $\mathbf{St}_G(v)$ es un grupo cíclico generado por la rotación en v de ángulo $2\pi/n$, que llamaremos f de aquí en adelante. La primera posibilidad es que $G = \mathbf{St}_G(v)$, es decir que G sea un **grupo cíclico de orden** n, cuyo único otro punto singular es $-v$; G consiste de rotaciones en el eje $\langle v, -v \rangle$. Si no es éste el caso, cualquier elemento de $G - \mathbf{St}_G(v) \neq \emptyset$ mueve a v.

Supongamos entonces que la órbita de $\boldsymbol{v}$ tiene algún punto distinto de $\boldsymbol{v}$. Como $G\,\boldsymbol{v}$ es finito podemos escoger $\boldsymbol{v}' \in G\,\boldsymbol{v}$ tal que $\boldsymbol{v} \neq \boldsymbol{v}'$ y que su distancia a $\boldsymbol{v}$ sea mínima. Por la demostración del Teorema del Gol (5.14), los centros de las rotaciones que mandan a $\boldsymbol{v}$ en $\boldsymbol{v}'$ estan en su mediatriz, ξ digamos; y como en G tenemos rotaciones que mandan a $\boldsymbol{v}$ en $\boldsymbol{v}'$, en ξ tenemos a sus correspondientes centros: sea $\boldsymbol{u}$ el más cercano de estos a $\boldsymbol{v}$. Entonces $\boldsymbol{u}$ es el centro de una rotación $g \in G$ tal que $g(\boldsymbol{v}) = \boldsymbol{v}'$.

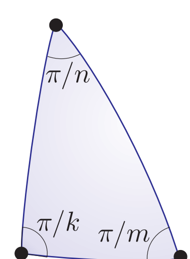
Como $\boldsymbol{u}$ es punto singular de G entonces su estabilizador es cíclico, de orden m digamos. Afirmamos que g es su generador de ángulo $\pm(2\pi/m)$. Si g no fuera este generador de $\mathbf{St}_G(\boldsymbol{u})$, entonces tendríamos que en $\mathbf{St}_G(\boldsymbol{u})$ hay una rotación g' de ángulo estrictamente menor que el de g (un divisor) y ésta mandaría a $\boldsymbol{v}$ en un punto más cercano que $\boldsymbol{v}'$ contradiciendo la forma en que escogimos a $\boldsymbol{v}'$. Así que g es la rotación en $\boldsymbol{u}$ de ángulo $\pm(2\pi/m)$; supongamos sin pérdida de generalidad (habría que cambiar $\boldsymbol{v}$ por $\boldsymbol{v}'$ si no) que es $2\pi/m$. Tenemos entonces que el ángulo de la línea $\boldsymbol{uv}$ a ξ es π/m.



Sea $\boldsymbol{w}$ el punto en ξ, tal que el triángulo $T = (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$ tiene ángulos internos π/m y π/n en $\boldsymbol{u}$ y $\boldsymbol{v}$ respectivamente; y sea γ el ángulo (dirigido) de T en $\boldsymbol{w}$. Por el Lema 5.15, $g \circ f$ es una rotación con centro en $\boldsymbol{w}$ y ángulo -2γ, que manda entonces a $\boldsymbol{v}$ en $\boldsymbol{v}'$. Como $g \circ f \in G$, se tiene que $\boldsymbol{w}$ es punto singular de G, y $g \circ f$ es generador de $\mathbf{St}_G(\boldsymbol{w})$ por la misma razón que g es generador de $\mathbf{St}_G(\boldsymbol{v})$ (que sólo usó cómo escogimos $\boldsymbol{v}'$). Tenemos entonces que $g \circ f$ es una rotación en $\boldsymbol{w}$ de un ángulo $\pm(2\pi/k)$, donde k es el orden de $\mathbf{St}_G(\boldsymbol{w})$.

Pero necesitamos demostrar aún más: que $\gamma = \pi/k$; y para esto hay que usar cómo escogimos $\boldsymbol{u}$. Hasta ahora sólo sabemos que 2γ, como ángulo, es $\pm(2\pi/k)$; y esto también es cierto si $\gamma = \pi \pm \pi/k$. Pero observemos que conforme un punto $\boldsymbol{x}$ en la mediatriz ξ se aleja de $\boldsymbol{v}$, partiendo del punto medio, el ángulo de la rotación con centro en $\boldsymbol{x}$ que lleva a $\boldsymbol{v}$ en $\boldsymbol{v}'$ decrece en valor absoluto; así que $\pi/m \geq \gamma$ pues $\boldsymbol{u}$ fue escogido como el más cercano a $\boldsymbol{v}$, es decir $d_{\mathbb{S}^2}(\boldsymbol{u}, \boldsymbol{v}) \geq d_{\mathbb{S}^2}(\boldsymbol{w}, \boldsymbol{v})$. Con este ingrediente extra, se puede concluir que $\gamma = \pi/k$.

Ahora sí, hemos encontrado un triángulo T cuyos vértices son centros de simetría rotacional de G de orden k, m, n y con ángulos internos correspondientes $\pi/k, \pi/m, \pi/n$. La clasificación de los grupos G que nos faltan depende de cuáles triángulos de estos existen.



Por el Teorema 5.10 (de la suma de ángulos de un triángulo esférico), se tiene que k, m, n (los tres enteros mayores que 1) deben cumplir

$$\frac{1}{k} + \frac{1}{m} + \frac{1}{n} > 1, \tag{5.8}$$

donde hemos factorizado y anulado a π. Si suponemos que $k \leq m \leq n$, es fácil ver que las únicas soluciones de esta ecuación son, en orden lexicográfico,

k	m	n	$\mathcal{A}(T)/\pi$	$\sharp G$
2	2	n	1/n	2n
2	3	3	1/6	12
2	3	4	1/12	24
2	3	5	1/30	60

(5.9)

donde se ha añadido la columna de exceso sobre 1 en (5.8), que, por el Teorema 5.10, multiplicada por π nos da el área del triángulo; y otra columna adicional donde, adelantándonos, ponemos la cardinalidad del grupo G de donde salió el triángulo, y que, como veremos, corresponde a $\mathcal{A}\left(\mathbb{S}^2\right)/2\mathcal{A}(T)$.

Hemos demostrado la parte de "unicidad" de un teorema —"si $G \subset \mathbf{SO}(3)$ es finito entonces cumple tal cosa"— y nos falta la de "existencia": ver que tales grupos existen. Pero antes de hacerlo vale un comentario sobre el final de nuestra argumentación previa. En el "es fácil ver que las únicas soluciones..." el argumento es que en la asignación que (lexicograficamente) sigue, la $2, 3, 6$, se obtiene la igualdad con 1 (en la desigualdad (5.8)); las otras dos igualdades con 1 son la $2, 4, 4$ y la $3, 3, 3$; para el resto de los valores de $k \leq m \leq n$, la suma de sus inversos es menor que 1. Las tres igualdades con 1 dan lugar a triángulos euclidianos y a tres grupos cristalográficos ahí, que son precisamente los subgrupos orientados (de isometrías que preservan orientación) de los tres grupos caleidoscópicos $K_{6,3,2}$, $K_{4,4,2}$ y $K_{3,3,3}$ que vimos en la Sección 3.9.2 (Teorema 3.27; de hecho, su demostración fué con la misma idea de encontrar un triángulo "racional"). El resto de las asignaciones, $k \leq m \leq n$ cuyos inversos suman menos que 1, daran lugar a subgrupos discretos del plano hiperbólico que veremos al final del capítulo 8; todos juntos son los llamados *grupos triangulares.* Dibujo...

Ahora sí, veamos la existencia de los grupos finitos $G \subset \mathbf{SO}(3)$ que, en abstracto, dieron lugar a la tabla (5.9). Para el primero, $22n$ (léase "dos-dos-ene"), pensemos en una pelota playera partida en n gajos iguales (angularmente de tamaño $2\pi/n$) con la válvula, donde confluyen los gajos, en el polo norte. Ahí está el centro de simetría rotacional de orden n. Los de orden 2 están todos en el ecuador. Una órbita en su intersección con las costuras y la otra en los puntos medios de los gajos. Al rotar π en uno de los primeros, un gajo se intercambia con el de junto y los polos también; al rotar π en uno de los de enmedio, su gajo intercambia hemisferios. Si pensamos en el plano del ecuador, éste siempre regresa a su lugar y aunque en $\mathbb{R}^3$ hayan sido rotaciones, las de orden 2 son reflexiones en él. Ese plano "cree" que el que actúa es el grupo diédrico de orden n, y efectivamente, este grupo es isomorfo a $\mathbf{D}_n$ pero realizado con rotaciones en $\mathbb{R}^3$. Dibujo...

Pasemos a exhibir al grupo 235, pues, esféricamente, es el que más manoseamos en estas épocas. Es el grupo de las simetrías orientadas del balón de fútbol: los movimientos (reales, físicos) que dejan su estructura de gajitos pentagonales y hexa-

gonales en su lugar (abstrayéndonos de la válvula y los letreros). El centro de simetría rotacional de orden 5, que, para fijar ideas, podemos suponer que es el tal $\boldsymbol{v}$ con el que arrancamos la demostración, es el centro de un pentágono: rotar $2\pi/5$ ahí, claramente, preserva la estructura. Como centro de orden 3 (que ha de ser $\boldsymbol{w}$) tomemos el centro de un hexagonito adyacente y obsérvese que su simetría es de orden 3 (no 6) pues a su alrededor alternan tres pentágonos y tres hexágonos. El de orden 2 (que apareció como $\boldsymbol{u}$ en la demostración) es entonces el punto medio de la costura entre el hexagonito escogido y un hexágono adyacente (también pegado al pentágono de $\boldsymbol{v}$); al rotar π ahí, se voltea la costura, se intercambian sus dos hexágonos y el pentágono original pasa al opuesto siguiendo la costura, cuyo centro es el tal $\boldsymbol{v}'$.

Es claro entonces que las órbitas de los tres tipos de puntos singulares son los centros de los pentágonos (orden 5) que hay 12, los de los hexágonos (orden 3) que hay 20 y los de las costuras entre estos (orden 2) que hay 30; y de cualquiera de ellos se obtiene que el grupo tiene 60 elementos. Obsérvese también que en la demostración pudimos haber empezado con cualquiera de las órbitas y haber llegado al mismo tipo de triángulo T, pero el punto auxiliar $\boldsymbol{v}'$, dependiendo de cómo fue que arrancamos, hubiera sido el nuevo que se obtiene al reflejar a T en sus tres lados (¿en qué orden hubieran aparecido los puntos en la demostración de haber empezado con uno de orden 3 o 2?). Al mover el cuadrilátero que forman T junto con alguno de sus reflejados por el grupo, se cubre toda la esfera sin traslapes (salvo en aristas o vértices), de aquí la relación entre las dos últimas columnas de la tabla. En fin, la existencia del grupo 235, la demuestra cualquier balón de fútbol aunque clásicamente se usan como referencia el *dodecaedro* y el *icosaedro.*

Dibujos...

El *dodecaedro* tiene como vértices los centros de simetría de orden 3; se obtiene de ellos tomando su *casco convexo*: todo lo que atrapan o está entre ellos y que formalmente se define como el conjunto de todas sus posibles combinaciones afines con coeficientes positivos. Su frontera está formada de doce (y de ahí el nombre) caras pentagonales que se juntan de a tres por vértice. Análogamente, o de manera dual, el *icosaedro* es el casco convexo de los centros de simetría de orden 5, que queda formado por caras triangulares de a cinco por vértice. En ambos casos, los centros de simetría de orden 2 están en los puntos medios de las aristas.

El grupo 234 es el de las simetrías orientadas del cubo (del dado borrado o, esféricamente, del balón de vólibol), cuyo dual, con el que comparte el grupo, es el octaedro (ocho triángulos de a cuatro por vértice). Por último, el 233 es el del tetraedro (cuatro triángulos equilateros pegados) que es autodual. Obsérvese que en este grupo, aunque sólo hay puntos singulares de orden 2 y 3, estos últimos se dividen en dos órbitas que corresponden a los cuatro vértices y a los centros de las caras.

Habiendo expuesto los posibles grupos, podemos resumir.

Teorema 5.17 *Los subgrupos finitos de* $\boldsymbol{SO}(3)$ *son los cíclicos* $\mathbf{C}_n$*, los diédricos orientados* $\mathbf{D}_n$ *(generados por dos medias vueltas) y los tres grupos de simetrías orientadas de los sólidos platónicos.* □

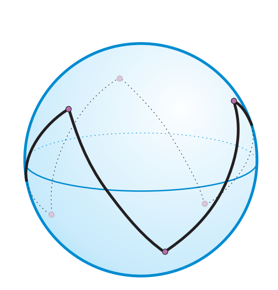
EJERCICIO 5.42 ¿Cuál es el grupo de simetrías orientadas de la pelota de tenis?

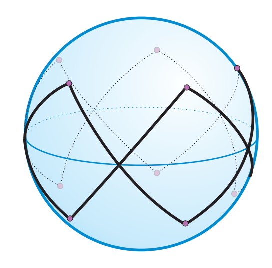
5.4.2 Subgrupos finitos no orientados

Veremos ahora qué posibilidades tiene $G \subset \mathbf{O}(3)$ finito si además tiene algún elemento que invierte orientación (y por lo tanto la mitad de ellos lo hacen).

Consideremos primero el caso general de un paso ρ en una línea ξ. ¿Qué genera? Si ρ está dentro de un grupo finito, al considerar su acción dentro de la línea ξ, que es un círculo, ρ genera ahí un grupo cíclico y podemos suponer que entonces tiene un ángulo $2\pi/n$ (si no, lo cambiamos por su potencia que sí lo tiene). Para fijar ideas y tener una figura geométrica en mente, supongamos que ξ es el ecuador y tomemos un punto $\boldsymbol{v}_0$ en el hemisferio norte (distinto del polo); $\rho(\boldsymbol{v}_0) =: \boldsymbol{v}_1$ queda en el hemisferio sur. Si tomamos el segmento de $\boldsymbol{v}_0$ a $\boldsymbol{v}_1$ y le aplicamos ρ, nos queda un segmento que sale de $\boldsymbol{v}_1$, vuelve a cruzar el ecuador hacia el norte y llega a $\boldsymbol{v}_2 := \rho(\boldsymbol{v}_1) = \rho^2(\boldsymbol{v}_0)$, y podemos seguir generando un zigzag con $\boldsymbol{v}_i := \rho^i(\boldsymbol{v}_0)$ que "zurce" el ecuador con segmentos de $\boldsymbol{v}_i$ a $\boldsymbol{v}_{i+1}$. Es un polígono espacial o esférico que llamaremos Z_n, "el zigzag-n".

Hay de dos sopas, si n es impar, $\boldsymbol{v}_n$ queda en el hemisferio sur; pues ρ^n regresa al círculo ξ a su lugar pero refleja sus dos lados, es justo la reflexión en ξ, ya que al componer un número impar de transformaciones que invierten orientación, ésta se invierte. Por la misma razón, si n es par, entonces $\rho^n = \mathrm{id}$, $\boldsymbol{v}_0 = \boldsymbol{v}_n$ y ρ genera un grupo cíclico de orden n que tiene como únicos puntos singulares a los polos y en ellos el estabilizador es cíclico de orden $n/2$ (con generador ρ^2), pues sólo en las composiciones pares estos puntos regresan a su lugar (lo que corresponde a que los $\boldsymbol{v}_i$ con i par están en el hemisferio norte.) Volviendo al caso en que n es impar (y entonces ρ^n es la reflexión en ξ y $\boldsymbol{v}_n$ el reflejado de $\boldsymbol{v}_0$), hay que dar una segunda vuelta para que $\boldsymbol{v}_0$ y también toda la esfera regrese a su lugar, $\rho^{2n} = \mathrm{id}$. En este caso, ρ genera un grupo cíclico de orden $2n$ que también tiene dos puntos singulares (los polos) con estabilizador de la mitad ($\mathbf{C}_n$ con generador de ángulo mínimo ρ^{n+1}), pero que además tiene puntos singulares en todo ξ, que llamaremos un *espejo*, cuyo estabilizador es la reflexión ρ^n (junto con id, por supuesto). Podemos resumir que los grupos finitos generados por un paso siempre son cíclicos de orden par, pero que ellos aparecen de dos formas distinguibles por la configuración de sus puntos singulares.





Hay una excepción en lo que hemos dicho; para $n = 1, 2$ no aparecen los centros de simetría rotacional en los polos. Para $n = 1$ tenemos una simple reflexión, y para $n = 2$, ρ es la función antípoda a que no tiene puntos ni rectas distinguidas: todos

sus estabilizadores son triviales (dícese que el grupo $\{id, a\}$ "*actúa libremente*" en $\mathbb{S}^2$); ξ sólo sirvió de apoyo pero a ella no la distingue el grupo.

...Dibujo,Z_n

Para $n > 2$, el zigzag-n, Z_n, es un ciclo o polígono combinatorio (de longitud n para n par y de longitud $2n$ para n impar), donde ρ actúa como su rotación generadora. Podemos agrandar al grupo $\langle\rho\rangle$ (el generado por ρ) para que tenga todas las posibles simetrías combinatorias del polígono Z_n, si le adjuntamos la reflexión σ en la línea perpendicular a ξ que pasa por v_0, y que combinatoriamente corresponde a la reflexión en el vértice básico. Entonces el grupo crece a un diédrico, donde, así como unas "rotaciones combinatorias" invertían orientación en la esfera (los pasos), las "reflexiones combinatorias" en las aristas son ahora rotaciones de π en las intersecciones de las aristas con el ecuador (la que "refleja" en el centro de la primera arista es $\sigma \circ \rho^{-1}$). Pero de nuevo, la configuración de puntos singulares distingue estos dos diédricos. Para n par, tenemos dos centros *caleidoscópicos* (con estabilizador diédrico) de orden $n/2$ en los polos junto con los $n/2$ espejos entre ellos y otros $n/2$ puntos de simetría rotacional de orden 2 en el ecuador (donde se intersectan las aristas de Z_n al ecuador y justo a medio camino entre los espejos). Pero para n impar, como ya teníamos un espejo en el ecuador, junto con las nuevas rotaciones de π ahí, se producen n centros caleidoscópicos de orden 2 (una órbita en los cruces dobles de aristas de Z_n y la otra entre pares de vértices opuestos de Z_n).

...Dibujo

Nos falta describir los *grupos caleidoscópicos*: finitos y generados por reflexiones. Los generados por dos reflexiones son diédricos (que ya aparecieron como subgrupos estabilizadores) y que esencialmente vienen de una dimensión inferior. Pero tenemos además los generados por tres reflexiones con espejos en los lados de un triángulo T "racional" con ángulos $\pi/k, \pi/m, \pi/n$ como en la Tabla 5.9 cuyo subgrupo que preserva orientación, es uno de los que ya estudiamos. Su configuración de puntos singulares es la misma que la de sus grupos de orientación, pero el estabilizador en cada punto se vuelve diédrico en vez de cíclico; y hay que añadir todos los espejos que parten a $\mathbb{S}^2$ en triangulitos equivalentes a T, tantos como elementos del grupo y justo el doble de lo indicado en la tabla. Obsérvese que de los primeros, la familia $22n$, los de n impar ya aparecieron como otra representación de los diédricos; pero para n par son nuevos.

Los tres que nos quedan son los grupos de simetrías de los sólidos platónicos; llamémoslos K_{233}, K_{234} y K_{235} que corresponden al tetraedro, al cubo (y octaedro) y al dodecaedro (e icosaedro). Se pueden obtener físicamente al poner tres espejos en los ángulos adecuados (claro, hay que cortar los espejos con cuidado y sabiendo cómo, pues los vidrios no son tan fáciles de mover e intersectar como nuestros planos abstractos en $\mathbb{R}^3$). Para obtener estos tres planos se toman un vértice v_0, el punto medio de una arista incidente v_1 y el centro, v_2, de una de las dos caras de la arista; los planos que estos puntos generan por parejas junto con el origen (el baricentro del solido platónico en cuestión) son los tres planos espejo del caleidoscopio correspondiente. Al asomarse a él se multiplica nuestra imagen, una por cada elemento del grupo, y viendo cada una a través de su triangulito. En la siguiente sección veremos con más cuidado

Foto o dibujo

cómo obtener estos puntos.

Hemos dado cuenta de todos los posibles grupos finitos de $\mathbf{O}(3)$. Falta demostrar que no hay más, la unicidad, pero lo dejamos como un posible trabajo o ejercicio avanzado.

EJERCICIO 5.43 ¿Cuáles son los subgrupos de orientación de $\mathrm{Sim}(Z_n)$?

EJERCICIO 5.44 ¿Cuántas reflexiones hay en K_{233} y cuántos pasos de qué ángulos? ¿Y en K_{234}?

5.4.3 Los grupos platónicos

En esta sección estudiamos con más cuidado los grupos de simetrías de los poliedros platónicos y damos representaciones explícitas de ellos.

Usando la notación introducida por Schläfli para poliedros regulares en el siglo XIX, denotemos por $\{\!\{m,n\}\!\}$ al poliedro platónico con m ágonos regulares por caras y con n de ellas en cada vértice.[1] Así $\{\!\{m,n\}\!\}$ puede ser $\{\!\{3,3\}\!\}$, $\{\!\{3,4\}\!\}$, $\{\!\{4,3\}\!\}$, $\{\!\{3,5\}\!\}$ o $\{\!\{5,3\}\!\}$ (tetraedro, octaedro, cubo, icosaedro, dodecaedro). Como en la sección anterior, llamemos $\mathbf{v}_0$,$\mathbf{v}_1$,$\mathbf{v}_2$ respectivamente a un vértice, el punto medio de una arista y el centro de una cara incidentes. Pensemos además que $\{\!\{m,n\}\!\}$ está centrado en el origen, y sea $\mathbf{u}_i$ el punto en $\mathbb{S}^2$ en la dirección $\mathbf{v}_i$, *i.e.*, $\mathbf{u}_i := (1/|\mathbf{v}_i|)\,\mathbf{v}_i$. Entonces los ángulos internos del triángulo esférico T con vértices $\mathbf{u}_0$,$\mathbf{u}_1$,$\mathbf{u}_2$ tiene ángulos respectivos π/n, $\pi/2$, π/m. Si llamamos ρ_i a la reflexión en el lado de T opuesto a $\mathbf{u}_i$, entonces ρ_i cumple que $\rho_i(\mathbf{u}_j) = \mathbf{u}_j \Leftrightarrow i \neq j$ (sólo mueve a $\mathbf{u}_i$). Pero además, como son reflexiones y puesto que $\rho_1\rho_0$ es la rotación generadora de la cara, $\rho_2\rho_1$ es la del vértice y $\rho_2\rho_0$ es la media vuelta en la arista, se tiene que cumplen las *relaciones de Coxeter*:

$$\begin{aligned} \rho_i^2 &= \mathbf{1} \quad \text{para} \quad i = 0,1,2 \\ (\rho_1\rho_0)^m &= (\rho_2\rho_1)^n = (\rho_2\rho_0)^2 = \mathbf{1}, \end{aligned} \tag{5.10}$$

donde entendemos por $\mathbf{1}$ a la identidad. Estas relaciones son importantes pues, en abstracto, definen por completo al grupo; se dice que dan una *presentación* de él por generadores y relaciones. No lo demostraremos, pero nótese que el análogo en dimensión 2 de los poliedros platónicos son los polígonos regulares. Si llamamos $\{\!\{m\}\!\}$ al de m lados, su grupo de simetrías está generado por dos reflexiones, ρ_0 y ρ_1, que cumplen las relaciones ...Dibujos

$$\rho_0^2 = \rho_1^2 = (\rho_1\rho_0)^m = \mathbf{1}.$$

Que esto define en abstracto el grupo diédrico $\mathbf{D}_m$ es porque al considerar ρ_0 y ρ_1 como letras y tomando en cuenta la equivalencia que determinan estas relaciones,

---

[1] Usaremos dobles corchetes para que no se confundan con conjuntos, aunque originalmente Schläfli usaba sólo uno.

las palabras en ρ_0, ρ_1 y sus inversos corresponden a los elementos de $\mathbf{D}_m$; cualquier palabra es equivalente (usando las sustituciones adecuadas en subpalabras) a una de tamaño chico y en la que se alternan las letras. No entramos en detalles, sólo queremos indicar que se abren puertas a generalizaciones naturales a más dimensiones y lo dejamos como motivación para un curso de álgebra y el estudio abstracto de grupos.

Dibujo...(cubo)

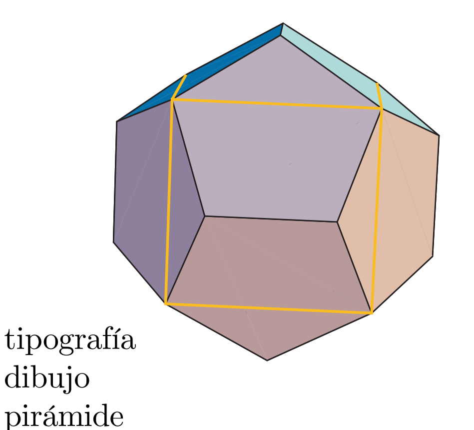
Lo que sí haremos es dar una *representación* explícita con matrices, es decir, dando una posición fija al poliedro $\{\!\{m, n\}\!\}$ en $\mathbb{R}^3$ y fijando las coordenadas de la terna básica $\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2$; estas reflexiones se expresarán como matrices que cumplen las seis relaciones de Coxeter (5.10).

Empecemos con el cubo $\{\!\{4, 3\}\!\}$. Sus coordenadas más elegantes se obtienen al tomar como vértices los ocho puntos de coordenadas $(\pm 1, \pm 1, \pm 1)$; y entonces es un cubo paralelo a los ejes y con arista de tamaño 2. Si tomamos como terna básica $\boldsymbol{v}_0 = (1, 1, 1)$, $\boldsymbol{v}_1 = (1, 0, 1)$, $\boldsymbol{v}_2 = (0, 0, 1)$, entonces la representación del grupo de simetrías del cubo, que habíamos llamado K_{234}, queda determinada por las matrices

$$\rho_0 \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \rho_1 \mapsto \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \rho_2 \mapsto \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

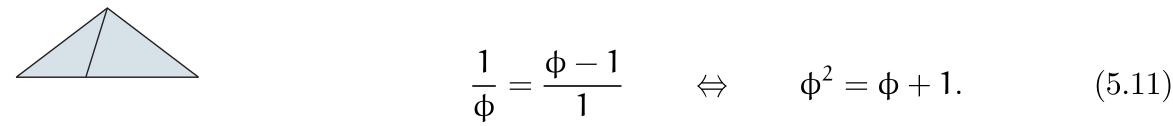
Se tiene entonces que $\rho_1\rho_0$ es la rotación de $\pi/2$ en el eje z, y que $\rho_2\rho_1$ es la matriz que permuta a la base canónica cíclicamente $\boldsymbol{e}_1 \mapsto \boldsymbol{e}_2 \mapsto \boldsymbol{e}_3 \mapsto \boldsymbol{e}_1$. Por lo que hemos visto, todas las multiplicaciones posibles de estas tres matrices dan 48 (cada columna y cada renglon con dos ceros y un ± 1) y son todas las simetrías del cubo.

El tetraedro, que conviene verlo inscrito en este cubo, se deja como ejercicio. Es más arduo establecer las coordenadas del dodecaedro $\{\!\{5, 3\}\!\}$. Conviene verlo directo hacia una arista; se ven entonces sólo cuatro caras, las dos pegadas por la arista y las dos "opuestas" a ella en sus vértices. Otras cuatro caras quedan de canto y las cuatro restantes atrás. Se observa entonces que los cuatro vértices que rodean a nuestra arista forman un cuadrado cuyo lado es la diagonal del pentágono, y que este cuadrado se completa hacia atrás para formar un cubo. Los vértices de este cubo son vértices del dodecaedro (se dice que está *inscrito*) y sus aristas son diagonales de las caras pentagonales, quedando cada cara del dodecaedro con una diagonal distinguida.



tipografía
dibujo
pirámide

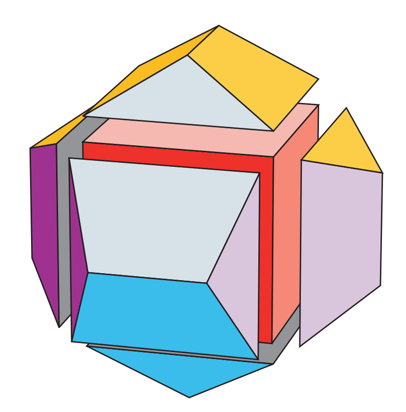
Tenemos que hacer un paréntesis y ver que la diagonal de un pentágono regular está en proporción áurea con su lado. Para esto, tomemos el triángulo isósceles en que corta una diagonal al pentágono. Éste es el isósceles *áureo chaparro*. Es fácil ver que sus ángulos miden $\pi/5$, $\pi/5$ y $3\pi/5$. Entonces al tomar un triangulito semejante con su base en el lado chico, lo que queda vuelve a ser isósceles, pues sus ángulos son $2\pi/5$, $2\pi/5$ y $\pi/5$: es el *áureo alto*. De aquí se sigue que sabemos cuánto mide el lado pequeño del semejante. Si el lado del pentágono mide 1, su diagonal ϕ debe cumplir por semejanza de triángulos que

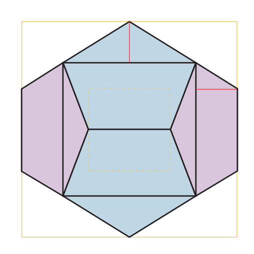


$$\frac{1}{\phi} = \frac{\phi - 1}{1} \qquad \Leftrightarrow \qquad \phi^2 = \phi + 1. \tag{5.11}$$

La solución positiva de esta ecuación es la proporción áurea (veéase Sección 3.9.3)

Cosideremos entonces un dodecaedro con arista de tamaño 2, de tal manera que ocho de sus vértices queden en los del cubo estándar de lado 2ϕ —es decir, son los ocho puntos $(\pm\phi, \pm\phi, \pm\phi)$—; para acabar de determinarlo hacemos que la arista por la que lo atraviesa —en su punto medio— el eje x esté en dirección del eje y. El dodecaedro se obtiene de este cubo pegándole adecuadamente una palapita en cada cara. El techo de cada palapa está hecho de dos triángulos y dos cuadriláteros y se ajustan los unos de una con los otros de otra para formar las caras pentagonales. Nos falta determinar la altura de la palapa que deberíamos llamar *áurea*.





Para esto, obsérvese que en la proyección ortogonal del dodecaedro al plano yz se ven los planos de la planta y las dos fachadas de la palapa áurea, y entonces la altura aparece en dos triángulos semejantes como sus dos catetos; los otros catetos los conocemos, son ϕ arriba, y $\phi - 1 = \phi^{-1}$ al lado. Por (5.11) la altura tiene que ser 1. De aquí que el punto medio de la arista más cercana a nosotros es $(\phi^2, 0, 0)$, de nuevo por (5.11), y los 12 vértices que nos faltan son $(\pm\phi^2, \pm 1, 0)$, $(0, \pm\phi^2, \pm 1)$ y $(\pm 1, 0, \pm\phi^2)$.

Si tomamos como triángulo básico $\mathbf{v}_0 = (\phi^2, 1, 0)$, $\mathbf{v}_1 = (\phi^2, 0, 0)$ y $\mathbf{v}_2$ como el centro de la cara superior (no necesitaremos sus coordenadas), entonces dos de las reflexiones son muy fáciles:

$$\rho_0 \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad \rho_2 \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Para encontrar ρ_1, podemos ver a dónde tienen que ir los tres vectores canónicos $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$, que no es difícil, pues en sus direcciones quedaron puntos medios de aristas. Por ejemplo, $\boldsymbol{v}_1$ tiene que ir bajo ρ_1 a

$$\frac{1}{2}\left(\begin{pmatrix}\phi\\\phi\\\phi\end{pmatrix}+\begin{pmatrix}\phi^2\\1\\0\end{pmatrix}\right)=\frac{1}{2}\begin{pmatrix}\phi^2+\phi\\\phi^2\\\phi\end{pmatrix}=\frac{1}{2}\begin{pmatrix}\phi^3\\\phi^2\\\phi\end{pmatrix},$$

cuya norma es ϕ^2 (esto se ve usando repetidamente la ecuación áurea (5.11) que implica $\phi^n = \phi^{n-1} + \phi^{n-2}$ y que nos permite bajar los grados de las expresiones). Entonces, normalizando, la primera columna de la matriz debe ser $(1/2)\left(\phi, 1, \phi^{-1}\right)$. Análogamente se encuentran las imágenes de $\boldsymbol{e}_2$ y $\boldsymbol{e}_3$ para obtener

$$\rho_1 \mapsto \frac{1}{2}\begin{pmatrix}\phi & 1 & \phi^{-1}\\ 1 & -\phi^{-1} & -\phi\\ \phi^{-1} & -\phi & 1\end{pmatrix}.$$

EJERCICIO 5.45 Encuentra las simetrías del tetraedro

---

[1] Para leer más sobre poliedros y otros puntos de vista ver [4], [11] y [10].

Capítulo 6

Geometría proyectiva

6.1 Motivación

Formalmente hablando, Desargues descubrió la geometría proyectiva; fundamentó matemáticamente los métodos de la *perspectiva* que habían desarrollado los artistas del renacimiento. Pero su trabajo, publicado en 1639, pasó desapercibido por 200 años (a excepción de dos teoremas famosos que se publicaron en otro lado). Quizá está desatención histórica se deba a que su contemporaneo Descartes se llevó las palmas. No es sino hasta el siglo XIX, cuando la geometría proyectiva, así como la hiperbólica, se establecieron como algo real, ineludible y cierto dentro de las matemáticas. Pero este proceso no fue fácil. Había que relajar la idea de **El Espacio** y **El Plano** (así, con mayúsculas y negritas) como algo único y dado —“ahí vivimos, se pensaba; está dado por Dios, se creía desde los griegos”—, así que pluralizar la idea a “los espacios” y “las geometrías”, casi como pasar de un cómodo monoteísmo a un exótico politeísmo, requirió tiempo y argumentos más allá de la simple lógica. Quizá lo que acabó de enraizarlas fue encontrar un modelo analítico: dentro del mismo contexto de la geometría euclidiana se puede construir la proyectiva, y entonces, si se acepta la primera, no queda más remedio que aceptar la segunda. Y este proceso acabó con contundencia a principios del siglo XX. Einstein, basándose en los impresionantes desarrollos geométricos de los matemáticos durante el recién rebasado “siglo de las geometrías”, demostró que, a gran escala, el universo se modela mejor con las nuevas geometrías que con el viejo y rígido espacio euclidiano.

No es nuestro interés seguir la línea de desarrollo histórico, pero sí vale la pena hacer una pausa para, a manera de motivación, abrir la mente a otros posibles enfoques.

6.1.1 El quinto postulado

Como mencionamos en las primeras páginas de este libro, el quinto postulado de Euclides recibió mucha atención por parte de los matemáticos de 20 siglos. Se dudaba

de la necesidad de incluirlo como axioma, reflejando la idea de que “El Plano” era único y estaba divinamente dado. Su negación tiene dos posibilidades: que no haya paralelismo, es decir, que cada par de rectas se intersecta en un punto; o bien, que el paralelismo no sea único, que por un punto fuera de una línea pasen muchas que no la intersectan. Al incluir la primera negación como quinto axioma se obtiene la geometría proyectiva; y al incluir la segunda la hiperbólica. Ahora entendemos bien que estos sistemas axiomáticos (el de la geometría proyectiva y el de la geometría hiperbólica) se dan, se construyen, en modelos basados en axiomas más elementales; tal y como construimos el plano y los espacios euclidianos basados en los axiomas de los números reales. Pero a principios del siglo XIX, que algún matemático demostrara teoremas basado en estos sistemas axiomáticos exóticos no implicaba que un poco más adelante, que a la vuelta de la esquina, no llegaría a una contradicción. Era trabajo de alto riesgo, cuestión de creer o no creer, y los “exóticos” ganaron la partida.

Regresando a la teoría que en este libro hemos desarrollado, tenemos a la mano la evidencia (y siempre estuvo ahí) de que el Quinto se puede negar proyectivamente. Acabamos de hacer geometría en la esfera; definimos líneas, distancias y ángulos, y vimos que hay un grupo que la hace homogénea; todo esto basados en los axiomas de los números reales. Nótese que la esfera casi cumple con los axiomas proyectivos excepto por que las líneas (¡cualquier par!) se intersectan en dos puntos en vez de uno, pero estas parejas de intersección son muy especiales: sólo para ellas falla la unicidad de la línea por dos puntos. Entonces, si declaramos que cada pareja de antípodas es un solo punto (esta abstracción es la que cuesta trabajo), se obtiene el **plano proyectivo rígido**. Así, el plano proyectivo es algo abstracto (sus puntos son parejas especiales de puntos, los antípodas), y la esfera es algo concreto que podemos ver y que lo “cubre” dos veces. Este modelo abstracto del plano proyectivo cumple, ahora sí, con todos los axiomas proyectivos: las líneas siempre se intersectan en un solo punto (piénsese “pareja antípoda”) y por cada par de puntos (piénsese, par de “parejas antípodas”) pasa una única línea. Además, a este plano proyectivo lo podemos ver en porciones pequeñas: es igual a un pedazo chico de $\mathbb{S}^2$; sólo la imagen global se dificulta.

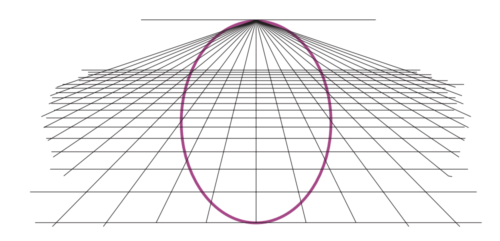
Sin embargo, no seguiremos la línea de desarrollo teórico que acabamos de esbozar. Habremos de regresar a ella, y entonces debe ser clara y sencilla, después de desarrollar el **plano proyectivo holgado**, que tiene muchó más simetrías porque no tiene noción de distancia. Podría decirse que es el análogo del plano afín (holgado) respecto al euclidiano (rígido). Y la motivación para entrarle viene de otro par de problemas o evidencias.

6.1.2 Las cónicas y el infinito

Mucho antes del surgimiento de la geometría analítica (Descartes), los griegos estudiaron las cónicas definidas como la intersección de un plano con un cono circular (¡y de ahí su nombre!). Ya las definimos, estudiamos y clasificamos (afín e isométricamente).

Aunque su definición sea única (insistimos, un cono intersectado con un plano, que no pasa por su ápice, precisamos), nos da tres tipos esencialmente distintos: elipses, parábolas e hipérbolas. Se pueden diferenciar por su comportamiento "al infinito": las elipses no tienen nada que ver con él; las parábolas se aproximan a él en una dirección importante para su propiedad focal, y las hipérbolas llegan al infinito en dos direcciones dadas por sus asíntotas, y se puede decir que pasan por él dos veces.

Más precisamente, la pendiente de las tangentes a la parábola tiende, cuando nos alejamos al infinito en cualquiera de sus dos direcciones, a la dirección de su eje, aunque la tangente tal cual se aleje al infinito. Podría decirse que la parábola es tangente al infinito, basta verla en perspectiva. Y de la hipérbola podemos decir aún más: si un punto en ella se aleja al infinito en cualquiera de sus cuatro posibilidades, su línea tangente, como línea, simple y plácidamente se acerca a una de las asíntotas y el movimiento natural de esa tangente es tocar a la asíntota, y seguirse como la tangente de puntos que se acercan por la rama opuesta. Como si la hipérbola simplemente cruzara al infinito por dos puntos correspondientes a sus dos asíntotas. (Véanse las animaciones del Capítulo 2 para sentir la naturalidad de lo que estamos diciendo.)



Debe haber una manera natural y coherente de incluir estos "puntos al infinito" para que las tres cónicas se conviertan en una sola cosa (como los griegos lo intuyeron). Consiste en agregarle al plano afín una recta al infinito cuyos puntos están asociados a las posibles direcciones de las rectas; y entonces las tres cónicas se completan y diferencian por su relación con esa recta, o no la tocan, o son tangentes o la cruzan en dos puntos (lo mismo que sucede con el resto de las rectas). Lograr que esa recta (que parece inalcanzable pero que se ve tan clara como el horizonte) se convierta en una recta como cualquier otra es lo que hace el plano proyectivo.

La técnica y la soltura matemática que se obtiene al formalizar lo anterior, permitió que la geometría algebraica se desarrollara en la segunda mitad del siglo XIX. Los inicios y la aplicación concreta a cónicas se verán en el siguiente capítulo. Por lo pronto, dejemos de divagar y concentrémonos en el problema que da la guía y la pauta del desarrollo formal de este capítulo. Además es el problema que dio su nombre a la geometría proyectiva y que históricamente tiene la primicia de su desarrollo.

6.1.3 El problema del pintor

...Dibujo: ¿Reproducci Durero?

En el Renacimiento se fue haciendo claro que un dibujo "realista" de una escena cotidiana se logra proyectando desde un "foco" (ojo hipotético) sobre el lienzo. Es decir, a cada punto, $\mathbf{x}$ digamos, de la escena le corresponde un punto en el lienzo, Λ digamos, al tomar la recta que pasa por el foco y $\mathbf{x}$ e intersectarla con Λ. La razón por la que los cuadros dibujados con esta técnica nos parece "realista" es que si se le observa desde el foco, la imagen que se percibe es casi idéntica (en teoría idéntica

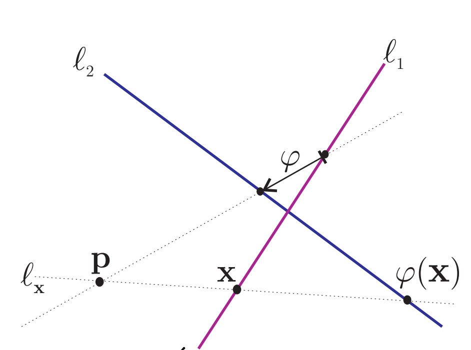
viendo con un solo ojo) a la escena original. Y si vemos el cuadro desde otro punto cualquiera, nuestro cerebro se encarga de adaptarlo para recrear la tridimensionalidad que se plasmó en el lienzo. Cómo logra esto nuestra mente, es harina de otro costal, pero que es un hecho que sabe hacerlo, y muy bien, se demuestra observando que de dos imágenes planas obtenidas con el mismo principio (las imágenes que se proyectan en la retina) tenemos una idea muy clara del espacio tridimensional que habitamos. Los pintores renacentistas, entre quienes destaca Durero, desarrollaron técnicas para dibujar con este principio. Las llamaron "métodos de la perspectiva" e incluyen el uso de "puntos de fuga". Como veremos más adelante estos puntos no son otra cosa que "puntos al infinito" que, al desarrollar la geometría proyectiva, se pueden tratar como puntos cualesquiera.

...Dibujo rieles

El problema concreto con el que empezamos la discusión formal en este capítulo puede plantearse en este contexto; es, por así decirlo, el problema que motivo a Desargues pero usando los métodos de su contemporáneo Descartes. Pensemos que queremos dibujar, en perspectiva, una vía de tren sobre una planicie inmensa. Los dos rieles se dibujan como segmentos que convergen a un punto de fuga. El problema es cómo dibujar los durmientes. Si en la realidad estan espaciados cada metro, digamos, ¿cada cuántos centímetros habrá que dibujarlos? Dicho de otra manera, ¿cómo cambia la escala al proyectar desde un punto?. Este problema, numérico y de tres a dos dimensiones, es el que resuelven automáticamente las computadoras en la realidad virtual. Veremos al final cómo lo hacen.

6.2 La línea proyectiva

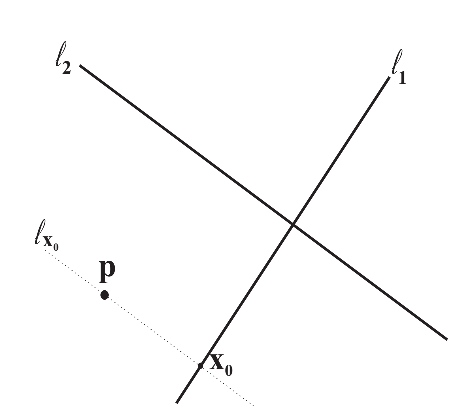
6.2.1 Proyecciones de rectas en rectas



Supongamos que tenemos dos rectas ℓ_1 y ℓ_2 en $\mathbb{R}^2$ parámetrizadas con parametros s y t respectivamente. Sea $\mathbf{p} \in \mathbb{R}^2$ un punto fuera de ellas. Podemos definir la *proyección desde* $\mathbf{p}$ de ℓ_1 a ℓ_2 geométricamente como sigue. Dado $\mathbf{x} \in \ell_1$, sea $\ell_\mathbf{x}$ la recta que pasa por $\mathbf{p}$ y $\mathbf{x}$; sea entonces

$$\text{“}\varphi : \ell_1 \to \ell_2\text{”} \qquad (6.1)$$
$$\varphi(\mathbf{x}) = \ell_2 \cap \ell_\mathbf{x} \,.$$

Ésta no es una función bien definida (de ahí que hayamos usado comillas) pues si $\mathbf{x}_0 \in \ell_1$ es tal que $\ell_{\mathbf{x}_0}$ es paralela a ℓ_2 entonces $\ell_2 \cap \ell_{\mathbf{x}_0} = \emptyset$ y $\varphi(\mathbf{x}_0)$ no tiene sentido. Pero este problema sólo se da en un punto, así que nos podemos hacer de la vista gorda y continuar. La pregunta que nos interesa responder es: ¿cómo se expresa φ en términos de los parámetros s y t?



Veamos primero un ejemplo concreto y sencillo, donde $\mathbf{p} = 0$,

$$\begin{aligned} \ell_1 &= \{(1,t) \mid t \in \mathbb{R}\}, \\ \ell_2 &= \{(s,1) \mid s \in \mathbb{R}\}. \end{aligned}$$

Dado $x = (1,t) \in \ell_1$, la recta ℓ_x es $\{(r, rt) \mid r \in \mathbb{R}\}$. Así que $\ell_x \cap \ell_2$ estará dado por la r que cumple $rt = 1$, es decir $r = 1/t$, y por lo tanto

$$\varphi(1,t) = (1/t, 1),$$

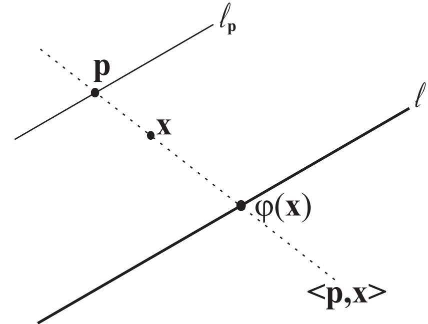
que está bien definida excepto cuando $t = 0$, lo que corresponde a que la recta por $(1,0)$ y el origen no intersecta a ℓ_2. Pensando a φ como función entre los parámetros, tenemos entonces

$$\text{“}\,\varphi(t) = \frac{1}{t}\,\text{”},$$

donde el entrecomillado significa que hemos abusado de la notación. Para hacer esto más preciso, y de una vez más general, notemos primero que la proyección desde un punto a una recta se puede extender a casi todo el plano.

Fijando una recta ℓ y un punto $\mathbf{p}$ fuera de ella, podemos definir para $x \in \mathbb{R}^2$, a $\varphi(x) \in \ell$ como la intersección de la recta que pasa por $\mathbf{p}$ y por x (que denotaremos $\langle \mathbf{p}, x\rangle$)[1] con la recta ℓ:

$$\varphi(x) = \langle \mathbf{p}, x\rangle \cap \ell.$$



Esta definición tiene sentido sólo cuando x no está en la paralela a ℓ por $\mathbf{p}$, pues entonces no hay tal intersección. Sea $\ell_{\mathbf{p}}$ la paralela a ℓ que pasa por $\mathbf{p}$, entonces φ está bien definida como función

$$\varphi : \mathbb{R}^2 - \ell_{\mathbf{p}} \to \ell.$$

Por ejemplo, en el caso en que ℓ es la recta $y = 1$ (ℓ_2 del ejemplo anterior) y que $\mathbf{p} = 0$, se tiene que

$$\varphi(x,y) = (x/y, 1)\,, \tag{6.2}$$

que está bien definida cuando $y \neq 0$.

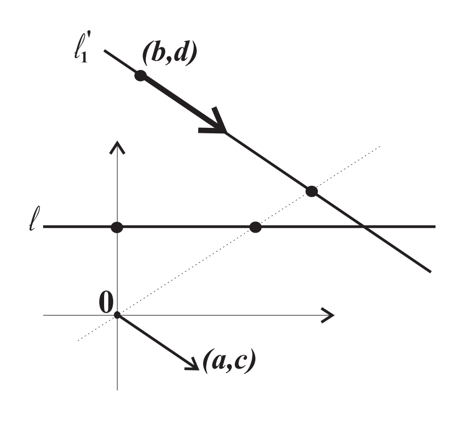
[1] Ya habíamos usado la notación $\langle -, - \rangle$ para el plano generado por dos vectores, pero por el contexto en que se usa no debe causar conflictos, y más adelante se justificará este abuso.

Ahora sí estamos en posición de demostrar el siguiente lema que responde a nuestra pregunta inicial.

Lema 6.1 *Sean* ℓ_1 *y* ℓ_2 *dos rectas en* $\mathbb{R}^2$ *parametrizadas afínmente con parámetros* s *y* t *respectivamente; y sea* $\mathbf{p}$ *un punto fuera de ellas (*i.e. $\mathbf{p} \notin \ell_1 \cup \ell_2$*). La proyección desde* $\mathbf{p}$ *de* ℓ_1 *a* ℓ_2 *tiene la expresión paramétrica*

$$\mu(s) = \frac{as + b}{cs + d}$$

para algunas constantes $a, b, c, d \in \mathbb{R}$ *y no todas ellas cero.*



Demostración. Por el Teorema ?? existe una única transformación afín de $\mathbb{R}^2$ que manda al punto $\mathbf{p}$ en el origen y a la recta ℓ_2 en la recta $\ell : y = 1$ haciendo corresponder al parámetro t con el parámetro natural x; es decir, que manda a $\mathbf{p}$ en 0, al punto base en ℓ_2 ($t = 0$) en $(0, 1)$ y al punto en ℓ_2 con $t = 1$ en $(1, 1)$. Esta transformación afín manda a la recta ℓ_1 en una nueva recta ℓ_1' que, al parametrizarse con s, tiene la expresión

$$\ell_1' = \{(as + b, cs + d) \mid s \in \mathbb{R}\},$$

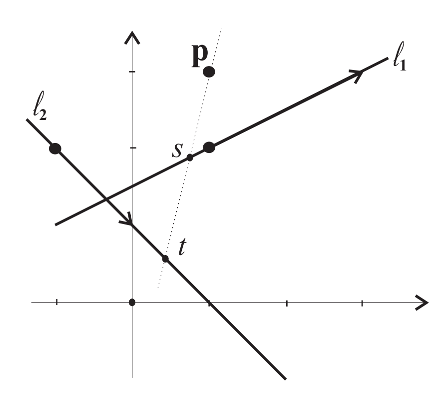
para algunas $a, b, c, d \in \mathbb{R}$ no todas cero, pues $(b, d) \neq (0, 0)$ ya que $\mathbf{p} \notin \ell_1$. Puesto que la proyección desde un punto es una construcción geométrica que no cambia por transformaciones afines (ya que éstas preservan rectas), entonces de la ecuación (6.2) y lo anterior se obtiene el lema. □

Para fijar ideas, veamos un ejemplo concreto y a través de él la idea de otra demostración que si siguieramos en general daría una enorme complicación de letras, aunque no así de números.

Ejemplo

Sean $\mathbf{p} = (1, 3)$,

$$\begin{aligned} \ell_1 &= \{(2s + 1, s + 2) \mid s \in \mathbb{R}\}, \\ \ell_2 &= \{(t - 1, 2 - t) \mid t \in \mathbb{R}\}. \end{aligned}$$



Queremos escribir la proyección φ de ℓ_1 a ℓ_2 desde $\mathbf{p}$ definida en (6.1) como una función μ entre los parámetros correspondientes. Para hacerlo, será conveniente escribir las rectas como funciones, es decir $\ell_1(s) = (2s+1, s+2)$ y análogamente para ℓ_2. Tenemos entonces que

$$\begin{aligned}
\mu(s) = t \quad & \Leftrightarrow \quad \ell_1(s), \ell_2(t) \text{ y } \mathbf{p} \text{ están alineados} \\
& \Leftrightarrow \quad \ell_1(s) - \mathbf{p} \parallel \ell_2(t) - \mathbf{p} \\
& \Leftrightarrow \quad (2s, s-1) \parallel (t-2, -t-1) \\
& \Leftrightarrow \quad 2s(-t-1) - (s-1)(t-2) = 0,
\end{aligned}$$

donde usamos el criterio de que dos vectores son paralelos si y sólo si su determinante es cero. Y de esta última ecuación ya se puede despejar t en términos de s. Al desarrollarla nos da:

$$\begin{aligned}
-3st + t - 2 &= 0 \\
t(-3s+1) &= 2 \\
t &= \frac{-2}{3s-1} \, .
\end{aligned}$$

que nos dice que la función buscada es

$$\mu(s) = \frac{-2}{3s-1} \, .$$

EJERCICIO 6.1 En el ejemplo anterior, encuentra la función inversa de μ, es decir la expresión paramétrica de la proyección desde $\mathbf{p}$ de ℓ_2 en ℓ_1; llamémosla ν.

EJERCICIO 6.2 ¿En dónde no están definidas las funciones μ y ν? ¿A qué corresponden geométricamente? Haz un dibujo.

EJERCICIO 6.3 Existe un único punto en ℓ_2 que no está en la imagen de la proyección. ¿A qué t corresponde, y cómo se obtiene geométricamente?

EJERCICIO 6.4 Análogamente, cuál es la única s que no está en la imagen de ν. Dibújala.

EJERCICIO 6.5 Sean $\ell_1 = \{(s-1, 2s+2) \mid s \in \mathbb{R}\}$ y $\ell_2 = \{(3t+2, -t+1) \mid t \in \mathbb{R}\}$. Encuentra las expresiones paramétricas de las proyecciones de ℓ_1 a ℓ_2 y de ℓ_2 a ℓ_1 desde los puntos $\mathbf{p} = (-1, 1)$ y $\mathbf{q} = (2, 2)$.

6.2.2 Transformaciones de Moebius

Motivados por la sección anterior definamos una *transformación de Moebius*[2] como la expresión

$$\mu(x) = \frac{ax+b}{cx+d}\ , \tag{6.3}$$

donde x es una variable en $\mathbb{R}$ y $a, b, c, d \in \mathbb{R}$ son constantes y no todas ellas son cero.

La hemos llamado "expresión" y no función pues tiene un problema. Para el número real x_0 que cumple $cx + d = 0$ ($x_0 = -d/c$) no está definida como número real, pues no se puede dividir por cero. Sin embargo, es claro que el límite de $\mu(x)$ cuando x tiende a x_0 es infinito. Al aproximarnos por un lado de x_0 (de tal manera que $cx + d$ tenga el mismo signo que $ax + d$) da valores tan grandes como queramos (nos aproximamos a ∞ por el lado positivo), y al aproximarnos a x_0 por el otro lado nos aproximamos a ∞ por el lado negativo. Si le añadimos a la recta real un punto al infinito, μ mandaría naturalmente a $-d/c$ al infinito. Formalicemos esta idea:

Definición 6.2.1 La recta proyectiva, $\mathbb{P}^1$, es

$$\mathbb{P}^1 := \mathbb{R} \cup \{\infty\},$$

es decir, consta de los números reales junto con un nuevo punto llamado infinito y denotado ∞ al que se aproxima uno creciendo tanto por el lado positivo de la recta como por el negativo.[3]

Y entonces hemos demostrado una parte del siguiente lema.

Lema 6.2 *La transformación de Moebius (6.3) define una función*

$$\mu : \mathbb{P}^1 \to \mathbb{P}^1$$

que es biyectiva si y sólo si $ad - bc \neq 0$.

Demostración. Hemos visto que μ está bien definida en todo $\mathbb{R}$, haciendo $\mu(-c/d) = \infty$. Nos falta definir $\mu(\infty)$; y esta definición de nueva cuenta es natural pues μ tiene un límite bien definido en ∞. Esto se ve cambiando de variable:

$$\lim_{x\to\infty} \frac{ax+b}{cx+d} = \lim_{y\to 0} \frac{a(1/y)+b}{c(1/y)+d} = \lim_{y\to 0} \frac{a+by}{c+dy} = \frac{a}{c}\ .$$

Definimos entonces $\mu(\infty) = a/c$ si $c \neq 0$, y $\mu(\infty) = \infty$ si $c = 0$. Por lo tanto μ está bien definida en todo el dominio.

---

[2] *real*, deberíamos especificar, pues el uso más común del término es con el campo de los números complejos $\mathbb{C}$ en vez del de los reales $\mathbb{R}$; nótese que sólo se usará que $\mathbb{R}$ es un campo.

[3] La recta proyectiva compleja, $\mathbb{CP}^1 := \mathbb{C} \cup \{\infty\}$, con las transformaciones de Moebius complejas producen una geometría riquísima (ver, e.g. [24]), también conocida como *esfera de Riemann*.

Para ver si es biyectiva o no, intentaremos encontrar la función inversa. Pensamos entonces en una nueva variable, y digamos, definida por

$$y = \frac{ax+b}{cx+d} \ ,$$

y tratemos de despejar x:

$$\begin{aligned} y(cx+d) &= ax+b \\ cxy+dy-ax &= b \\ x(cy-a) &= -dy+b, \end{aligned}$$

de donde

$$x = \frac{dy-b}{-cy+a} \ .$$

Hasta ahora, no hemos puesto cuidado en si dividimos por cero o no; hemos procedido intempestivamente. Pero procedamos ahora más formalmente para demostrar lo que nos falta. Esta última expresión nos dice cuál, si acaso, debía ser la inversa. Así que definamos

$$\nu(x) = \frac{dy-b}{-cy+a} \ . \tag{6.4}$$

Por la primera parte de esta demostración, ν es una función de $\mathbb{P}^1$ en $\mathbb{P}^1$; y cumple

$$\begin{aligned} (\nu \circ \mu)(x) &= \frac{d\left(\frac{ax+b}{cx+d}\right)-b}{-c\left(\frac{ax+b}{cx+d}\right)+a} \\ &= \frac{adx+bd-bcx-bd}{-acx-bc+acx+ad} \\ &= \frac{(ad-bc)x}{ad-bc} \ . \end{aligned}$$

Si $ad-bc \neq 0$, esta última expresión es x (para cualquier $x \in \mathbb{R}$ y también cuando $x = \infty$ usando las definiciones); de tal manera que μ es biyectiva pues su inversa es ν.

Nos falta ver que si $ad-bc = 0$ entonces μ no es biyectiva; de hecho es constante. Esto se puede ver por cuatro casos cuando alguna de las constantes es distinta de cero. Por ejemplo, si $d \neq 0$, multiplicando arriba y abajo por d en la expresión de μ se obtiene

$$\begin{aligned} \mu(x) &= \frac{d(ax+b)}{d(cx+d)} = \frac{adx+bd}{d(cx+d)} \\ &= \frac{bcx+bd}{d(cx+d)} = \frac{b(cx+d)}{d(cx+d)} = \frac{b}{d} \ , \end{aligned}$$

independientemente de la x. Como los otros casos, que se dejan al lector, son análogos, esto demuestra el lema. □

Observemos que en la demostración anterior encontramos la fórmula de la inversa de una función de Moebius y que ésta de nuevo resultó del mismo tipo. Si definimos a las *proyectividades de* $\mathbb{P}^1$ (también conocidas como las *transformaciones de Moebius*) como

$$\mathbf{Pr}(1) := \{\mu : \mathbb{P}^1 \to \mathbb{P}^1 \mid \mu(x) = \frac{ax+b}{cx+d};\ ad - bc \neq 0\}$$

tenemos entonces:

Teorema 6.3 ***Pr***(1) *es un grupo de transformaciones de la recta proyectiva* $\mathbb{P}^1$*.*

Demostración. Como la identidad es claramente una transformación de Moebius (tómese $a = d = 1$ y $b = c = 0$) y hemos visto que sus inversas también lo son, sólo nos falta ver que la composición de transformaciones de Moebius también es de Moebius. Sean

$$\mu(x) = \frac{ax+b}{cx+d} \quad \text{y} \quad \nu(x) = \frac{\alpha x+\beta}{\gamma x+\delta}\ .$$

Se tiene entonces que

$$\begin{aligned}(\mu\circ\nu)(x) &= \frac{a\left(\frac{\alpha x+\beta}{\gamma x+\delta}\right)+b}{c\left(\frac{\alpha x+\beta}{\gamma x+\delta}\right)+d} = \frac{a\,(\alpha x+\beta)+b(\gamma x+\delta)}{c\,(\alpha x+\beta)+d(\gamma x+\delta)} \\ &= \frac{(a\alpha+b\gamma)x+(a\beta+b\delta)}{(c\alpha+d\gamma)x+(c\beta+d\delta)} \in \mathbf{Pr}(1)\ . \qquad (6.5)\end{aligned}$$

□

Si pensamos que la recta proyectiva es una extensión de la recta real (se le añadió el punto al infinito), resulta entonces que las transformaciones afines son precisamente las proyectivas que fijan al infinito. Es decir, las proyectivas son una extensión de las afines. Pues si $\mu \in \mathbf{Pr}(1)$ es tal que $\mu(\infty) = \infty$, entonces (siguiendo con la expresión (6.3)), como $\mu(\infty) = a/c$, se tiene que $c = 0$. Esto implica que $d \neq 0$ (pues $ad - bc \neq 0$) y por tanto

$$\mu(x) = (a/d)x + (b/d),$$

que es una transformación afín. Hemos demostrado que $\mathbf{Af}(1) \subset \mathbf{Pr}(1)$, y más precisamente que

$$\mathbf{Af}(1) = \{\mu \in \mathbf{Pr}(1) \mid \mu(\infty) = \infty\},$$

pues estas transformaciones de Moebius pueden interpretarse como transformaciones de $\mathbb{R}$ en $\mathbb{R}$.

EJERCICIO 6.6 Para diversas transformaciones de Moebius (las que aparecen en los ejercicios anteriores, cuando menos), encuentra sus composiciones y sus inversas.

EJERCICIO 6.7 Sean ℓ_1 y ℓ_2 dos rectas parametrizadas y **p** un punto fuera de ellas. Demuestra que la transformación de Moebius que relaciona sus parámetros es afín si y sólo si las líneas son paralelas.

6.2.3 Teorema de 3 en 3

Si las transformaciones afínes mandan cualquier pareja de puntos en cualquier otra, ahora las proyectivas mandan ternas en ternas (correspondiendo al hecho de que con $\mathbf{Pr}(1)$ el infinito ya puede ir a cualquier otro punto).

Teorema 6.4 *Dadas dos ternas* $\alpha_0, \alpha_1, \alpha_2$ *y* $\beta_0, \beta_1, \beta_2$ *de puntos distintos en* $\mathbb{P}^1$*, existe una única transformación proyectiva* $\mu \in \boldsymbol{Pr}(1)$ *tal que* $\mu(\alpha_i) = \beta_i$ *para* $i = 0, 1, 2$*.*

Demostración. Veamos primero el caso en que $\alpha_0 = 0, \alpha_1 = 1, \alpha_2 = \infty$. Estamos buscando

$$\mu(x) = \frac{ax + b}{cx + d} \tag{6.6}$$

(*i.e.*, las constantes $a, b, c, d \in \mathbb{R}$) tal que $\mu(\alpha_i) = \beta_i$ para $i = 0, 1, 2$. Las condiciones nos dan tres ecuaciones

$$\frac{b}{d} = \beta_0\,, \quad \frac{a+b}{c+d} = \beta_1\,, \quad \frac{a}{c} = \beta_2\,. \tag{6.7}$$

Supongamos primero que las β_i son reales (*i.e.*, que ninguna es ∞). De la primera y la tercera ecuación se obtiene $b = \beta_0 d$ y $a = \beta_2 c$, que, al sustituir en la segunda, nos da

$$(\beta_2 - \beta_1)c + (\beta_0 - \beta_1)d = 0.$$

Esta ecuación lineal con dos incógnitas tiene como una solución no trivial (recuérdese el compadre ortogonal):

$$c = \beta_1 - \beta_0\,, \quad d = \beta_2 - \beta_1$$

y por tanto

$$a = \beta_2(\beta_1 - \beta_0)\,, \quad b = \beta_0(\beta_2 - \beta_1)\,.$$

Si alguna de las β_i fuera ∞ (digamos $\beta_0 = \infty$), la ecuación correspondiente nos indica que el denominador es cero (en el ejemplo, que $d = 0$). Y de cualquier manera el sistema tiene solución (que se deja al lector).

En el caso general (las α_i son cualesquiera), se puede proceder de dos maneras. La primera, más económica en casos particulares, sería traducir las ecuaciones $\mu(\alpha_i) = \beta_i$

a ecuaciones lineales para obtener el siguiente sistema de tres ecuaciones con cuatro incógnitas (a, b, c, d):

$$\begin{array}{ccccccccc} \alpha_0 a & + & b & - & \alpha_0\beta_0 c & - & \beta_0 d & = & 0 \\ \alpha_1 a & + & b & - & \alpha_1\beta_1 c & - & \beta_1 d & = & 0 \\ \alpha_2 a & + & b & - & \alpha_2\beta_2 c & - & \beta_2 d & = & 0 \end{array}$$

(en el caso en que todas las constantes sean reales, o uno equivalente cuando aparezcan infinitos); y luego encontrar una solución no trivial. O bien se puede proceder usando el caso que ya resolvimos y que las transformaciones de Moebius no constantes tienen inversa. Si ν (y, respectivamente, μ) es la transformación de Moebius que manda a $0, 1, \infty$ en $\alpha_0, \alpha_1, \alpha_2$ (y, respectivamente, en $\beta_0, \beta_1, \beta_2$), entonces la que estamos buscando es $\mu \circ \nu^{-1}$.

Falta demostrar que la transformación de Moebius que hemos encontrado es única. Para esto, consideremos primero un caso muy especial, cuando las dos ternas en cuestión son la *terna canónica* $0, 1, \infty$. Supongamos entonces que una transformacion de Moebius μ, dada por (6.6) cumple que $\mu(0) = 0$, $\mu(1) = 1$ y $\mu(\infty) = \infty$. Queremos demostrar que es la identidad. En este caso, las ecuaciones (6.7) nos dan:

$$\frac{b}{d} = 0\,, \quad \frac{a+b}{c+d} = 1\,, \quad \frac{a}{c} = \infty\,.$$

La primera dice que $b = 0$ y la tercera que $c = 0$, así que la segunda implica que $a = d$. Y en consecuencia

$$\mu(x) = \frac{a}{d}x = x,$$

como queríamos demostrar. Nótese que no se obtiene que $a = 1 = d$; simplemente demostramos que $a = d$ (y que son distintas de cero), y por tanto que $a/d = 1$. Pues aunque las transformaciones de Moebius están bien definidas, las constantes a, b, c, d que las definen no están específicamente determinadas: si multiplicamos a todas ellas por una misma constante no cero, obtenemos la misma función. El caso general se sigue de éste usando las propiedades de grupo. Dejamos los detalles al lector, pues a continuación lo veremos con una mejor notación. □

Matrices asociadas

Conviene recapitular un poco y establecer nueva notación. Una función de Moebius está determinada por cuatro constantes reales, que, como ya nos ha sucedido, y por varias "casualidades" que se nos han presentado, conviene empaquetar en una matriz. Sea A la matriz

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Si esta matriz es distinta de la matriz cero ($A \neq 0$), le podemos asociar la función de Moebius

$$\mu_A(x) = \frac{ax+b}{cx+d} \ .$$

Obsérvese que para $t \neq 0$ se tiene que $\mu_A = \mu_{tA}$, pues

$$\mu_{tA}(x) = \frac{tax+tb}{tcx+td} = \frac{t(ax+b)}{t(cx+d)} = \frac{ax+b}{cx+d}.$$

Esta asignación de matrices a funciones de Moebius tiene además las bonitas propiedades siguientes:

Primero, μ_A es una transformación (biyectiva) si y sólo si $\det(A) \neq 0$; es decir, si y sólo si A es invertible (veáse el Lema 6.2).

Segundo, en este caso, la inversa de la transformación de Moebius corresponde a la inversa de la matriz, pues en la ecuación (6.4) las constantes corresponden a la matriz $\det(A)A^{-1}$, que por lo anterior tiene asociada la función de Moebius $\mu_{A^{-1}}$ (podemos resumir, $(\mu_A)^{-1} = \mu_{A^{-1}}$).

Y **tercero**, por la ecuación (6.5), la composición de transformaciones de Moebius corresponde a la multiplicación de matrices, es decir, $\mu_A \circ \mu_B = \mu_{AB}$.

Estas tres propiedades se resumiran en la frase: "la función $\mathbf{GL}(2) \to \mathbf{Pr}(1)$ dada por $A \mapsto \mu_A$ es un *homomorfismo de grupos*".

Ahora sí podemos concluir con elegancia la unicidad en la demostración del Teorema de "tres en tres". Habíamos demostrado que si μ_A cumple que $\mu_A(0) = 0$, $\mu_A(1) = 1$ y $\mu_A(\infty) = \infty$, entonces A es diagonal ($a = d \neq 0$ y $b = c = 0$). Tenemos entonces que

$$\mu_A(0) = 0, \ \mu_A(1) = 1 \ \text{ y } \ \mu_A(\infty) = \infty \quad \Leftrightarrow \quad A = t\,I \quad \text{con} \quad t \neq 0.$$

Supongamos ahora que A es una matriz cuya transformación de Moebius fija tres puntos distintos cualesquiera; es decir, tal que $\mu_A(\alpha_i) = \alpha_i$ para $i = 0, 1, 2$ con $\alpha_0 \neq \alpha_1 \neq \alpha_2 \neq \alpha_0$. Por la primera parte del Teorema 6.4, existe una matriz invertible B tal que $\mu_B(0) = \alpha_0$, $\mu_B(1) = \alpha_1$ y $\mu_B(\infty) = \alpha_2$. Es fácil ver que entonces la transformación de Moebius $\mu_{B^{-1}} \circ \mu_A \circ \mu_B$ fija a $0, 1, \infty$ y, por el caso que ya vimos, existe $t \neq 0$ tal que

$$B^{-1}AB = t\,I.$$

Multiplicando por la izquierda por B y por la derecha por B^{-1} se obtiene

$$A = B(t\,I)B^{-1} = t\,(BIB^{-1}) = t\,(BB^{-1}) = t\,I\ .$$

Supongamos, por último, que A y B son dos matrices cuyas transformaciones de Moebius mandan una terna en otra terna igual, el caso general. Entonces $\mu_{A^{-1}} \circ \mu_B$ fija la primera terna, y por lo anterior se tiene que existe $t \neq 0$ tal que

$$A^{-1}B = t\,I,$$

multiplicando por A tenemos que

$$B = A(t\,I) = t\,(AI) = t\,A.$$

Lo cual demuestra que inducen la misma transformación de Moebius. Y queda demostrado que

$$\mu_A = \mu_B \quad \Leftrightarrow \quad B = t\,A \quad \text{con} \quad t \neq 0.$$

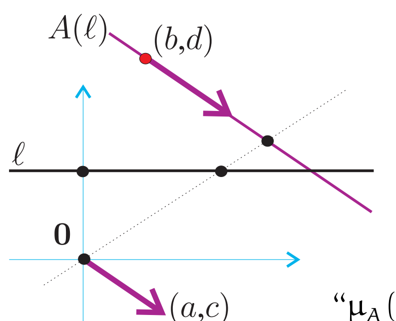
EJERCICIO 6.8 Considera las siguientes ternas de puntos en $\mathbb{P}^1$: $\mathbf{a} = (0, 1, \infty)$, $\mathbf{b} = (1, 2, 3)$, $\mathbf{c} = (-1, 0, 1)$, $\mathbf{d} = (-1, \infty, 1)$. Encuentra una expresión para las 12 transformaciones de Moebius que mandan, ordenadamente, a una de estas ternas en otra.

6.2.4 La recta proyectiva y $\mathbb{R}^2$

No por casualidad aparecieron las matrices para ayudarnos en la sección anterior. Recapitulemos. Las matrices invertibles habían surgido como los paquetes de información para definir las transformaciones lineales de $\mathbb{R}^2$. Y ahora surgen como paquetes de información para transformaciones de Moebius con la ambigüedad de una constante no cero. Para que la matriz A "actúe" en $\mathbb{R}^2$ teníamos la siguiente fórmula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}. \tag{6.8}$$

Recordemos también (6.2) que la proyección desde el origen de $\mathbb{R}^2$ a la recta canónica $y = 1$ tiene la expresión



$$\varphi \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x/y \\ 1 \end{pmatrix},$$

de tal manera que si identificamos al punto $(x, 1)$ en la recta canónica con su primera coordenada x en $\mathbb{R}$, podemos escribir (entrecomillado por el abuso de notación):

$$\text{``}\mu_A(x) = \varphi(A \begin{pmatrix} x \\ 1 \end{pmatrix}) = \varphi \begin{pmatrix} ax + b \\ cx + d \end{pmatrix} = \frac{ax + b}{cx + d}\text{''}.$$

Más formalmente, si ℓ es la recta canónica $y = 1$ y f es la función lineal asociada a A, podemos concluir que μ_A es la función de Moebius que relaciona los parámetros naturales de la recta $f(\ell) = \{(ax + b, cx + d) \mid x \in \mathbb{R}\}$ y la recta ℓ con la proyección desde el origen. Lo cual demuestra que toda transformación de Moebius surge de una proyección. Pero esto se puede formalizar de manera más elegante y sin entrecomillados.

Proporciones

Los antiguos griegos pensaban a los números reales como proporciones entre números enteros. Por ejemplo, lo que ahora conocemos como $1/2$ o 0.5, ellos lo pensaban como la proporción del 1 al 2, o del 2 al 4 o del 3 al 6... que en la escuela primaria se denota "$1 : 2 = 2 : 4 = ...$". Se llevaron una gran sorpresa cuando geométricamente encontraron un número (una distancia), a saber $\sqrt{2}$ (la hipotenusa del triángulo rectángulo isósceles con lado unitario), que no se podía expresar de esta manera. Ahora sabemos que no todos los números reales son racionales, pero en fin, la idea básica es buena y ahora la retomamos pero de manera más general, pensando en propociones no sólo de números enteros sino de números reales.

Cada número real $x \in \mathbb{R}$ se puede pensar como la proporción del x al 1, que es la misma proporción del $2x$ al 2, o del tx al t para cualquier t. Una pareja (x, y) cualquiera, como proporción $x : y$, representa al número que, multiplicado por y, nos da x. El 0 es entonces la proporción del 0 a cualquier t y podemos pensar al ∞ como la proporción de cualquier número al 0. En el texto anterior surge una terrible ambigüedad respecto a la pareja $(0, 0)$; vuélvase a leer con ese ejemplo maligno en mente y se verá que podría representar a cualquier proporción, a cualquier número. Pero si eliminamos a la pareja $(0, 0)$ y exigimos que t nunca sea cero, la ambigüedad desaparece. Formalicemos la idea, con la notación de las proporciones de primaria.

Definición 6.2.2 Dado $(x, y) \in \mathbb{R}^2$, $(x, y) \neq (0, 0)$, sea

$$[x : y] := \{(tx, ty) \mid t \in \mathbb{R},\ t \neq 0\}. \tag{6.9}$$

Que se puede leer "la proporción de x a y". Siempre que consideremos la expresión $[x : y]$, suponemos que no ambas "coordenadas" son cero.

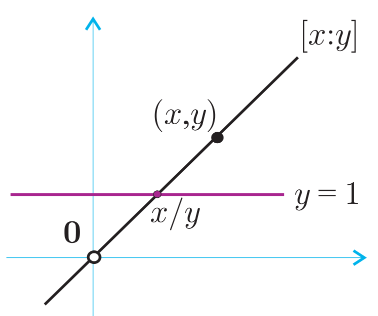
Observemos ahora que los puntos de la recta proyectiva corresponden biunivocamente a las proporciones, pues a $x \in \mathbb{R}$ le corresponde la proporción $[x : 1]$ y al ∞ le corresponde la proporción $[1 : 0]$. E inversamente, a la proporción $[x : y]$ le corresponde el número x/y (donde $x/0 = \infty$ cuando $x \neq 0$). De tal manera que podemos *redefinir a la recta proyectiva* como

$$\mathbb{P}^1 := \left\{[x : y] \mid (0, 0) \neq (x, y) \in \mathbb{R}^2\right\}.$$

A esta manera de ver la recta proyectiva se le conoce como la dada por *coordenadas homogéneas*. Así, un punto en $\mathbb{P}^1$ dado por coordenadas homogéneas es la proporción $[x : y]$. La gran ventaja es que ahora todos los puntos de $\mathbb{P}^1$ pueden ser tratados de la misma manera; ya no hay que hacer distinciones para ∞; la homogeneidad es evidente.

Rectas por el origen

Hemos definido a $[x : y]$ en (6.9) como un subconjunto de $\mathbb{R}^2$, que es precisamente una recta por el origen. Así que los puntos de $\mathbb{P}^1$ corresponden a las rectas por el origen en $\mathbb{R}^2$.



Si pensamos, como ya lo hicimos antes, en la recta $y = 1$ como los reales, la correspondencia $[x : y] \longmapsto x/y$ corresponde a intersectar la recta por el origen $[x : y]$ con la recta canónica $y = 1$, y la única que no tiene intersección (el eje de las x dado por $y = 0$) corresponde al infinito.

Conviene entonces introducir la notación

$$[\mathbf{x}] := [x : y]\,, \text{ donde } \ (0,0) \neq (x,y) = \mathbf{x} \in \mathbb{R}^2.$$

Esto nos da una función $(\mathbf{x} \mapsto [\mathbf{x}])$ de $\mathbb{R}^2 - \{\mathbf{0}\}$ sobre $\mathbb{P}^1$, en la cuál la imagen inversa de un punto en $\mathbb{P}^1$ es la recta por el origen a la que corresponde, pues obsérvese que por la definición $[\mathbf{x}] = [t\mathbf{x}]$ para cualquier $t \neq 0$.

Ahora, las transformaciones proyectivas se expresan en coordenadas homogéneas de manera muy natural. Dada una matriz $A \in \mathbf{GL}(2)$, su transformación proyectiva μ_A se expresa como

$$\mu_A : \mathbb{P}^1 \to \mathbb{P}^1$$
$$\mu_A([\mathbf{x}]) = [A\mathbf{x}]\,,$$

pues si A es como en (6.8) y $\mathbf{x} = (x, y)$ tenemos que

$$[A\mathbf{x}] = [ax + by : cx + dy]\,.$$

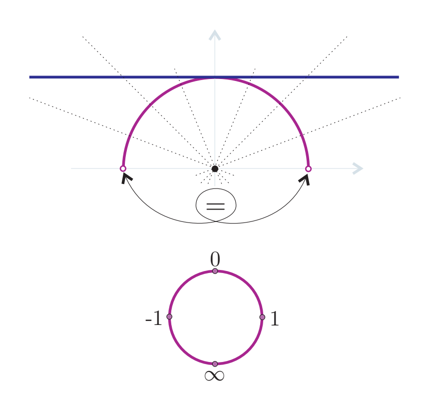
Esto concuerda con nuestra antigua idea de transformación de Moebius pues el número x en coordenadas homogéneas es $[x : 1]$ al cual A manda en $[ax + b : cx + d]$ que corresponde a $(ax + b)/(cx + d)$, y el infinito corresponde al $[1 : 0]$ que va en $[a : c]$ que corresponde a a/c como debía ser.

Este enfoque es el adecuado pues fácilmente se generaliza a más dimensiones. Pero antes de hacerlo haremos una clasificación topológica de las transformaciones de Moebius, y regresaremos brevemente con el pintor y algunos de sus problemas prácticos.

EJERCICIO 6.9 Con nuestra nueva definición de $\mathbb{P}^1$ y de transformaciones proyectivas, es decir, con coordenadas homogéneas y matrices, demuestra el Teorema de 3 en 3.

6.2.5 Tipos de transformaciones

Aunque no lo hayamos dicho explícitamente, a la recta proyectiva se le puede pensar también como un círculo. Puesto que a la recta real se le añade un único punto al que se llega por ambos extremos, éste cierra el círculo. O bien, pensado como rectas por el origen en $\mathbb{R}^2$, si tomamos el casquete superior del círculo unitario (definido por $\{(x,y) \in \mathbb{S}^1 \mid y \geq 0\}$), cada recta lo intersecta en un punto único excepto la horizontal que lo intersecta en los dos extremos; al identificarlos, nos queda un círculo (de la



mitad de la circunferencia). En esta sección, vamos a pensar a $\mathbb{P}^1$ como un círculo para describir topológicamente, es decir a grandes rasgos, cómo son las transformaciones de Moebius.

Los dos "invariantes" de las transformaciones de Moebius que nos servirán son el número de puntos fijos, y la orientación; es decir si preserva o no orientación. Esta última se detecta por el determinante de una matriz asociada; si es positivo, un punto viajando en $\mathbb{P}^1$ va en la misma dirección en que viaja su imágen, y si es negativo, éstos viajan en direcciones contrarias. Esto se ve claro en nuestro *modelo lineal* de $\mathbf{Pr}(1)$, es decir, pensando en líneas por 0 en $\mathbb{R}^2$. Pero observemos que el determinante no está definido para transformaciones de Moebius, pues sus matrices asociadas difieren por un factor no nulo; aunque su signo sí lo está, pues $\det(tA) = t^2 \det(A)$.

Veamos ahora los posibles puntos fijos de $\mu_A \in \mathbf{Pr}(1)$. Por el Teorema de 3 en 3, si fueran tres sería la identidad, así que $\mu_A \neq \mathrm{id}$ tiene a lo más dos puntos fijos. Esto también se puede ver de otras dos maneras. Primero, de la fórmula númerica al tratar de resolver la ecuación

$$x = \mu_A(x) = \frac{ax+b}{cx+d}$$

que se traduce en encontrar las raíces de la ecuación cuadrática

$$cx^2 + (d-a)x - b = 0.$$

Y segundo, buscando vectores propios (recordar el Capítulo 4) de la matriz A, pues son vectores $\mathbf{v}$ no nulos para los cuales $A\mathbf{v} = \lambda\mathbf{v}$ para alguna $\lambda \in \mathbb{R}$ (y por tanto $\mu_A[\mathbf{v}] = [A\mathbf{v}] = [\lambda\mathbf{v}] = [\mathbf{v}]$, pues $\lambda \neq 0$ porque $\det A \neq 0$). Con este enfoque, habría que encontrar los valores propios de A como las raíces de su polinomio característico

$$\det(A - \lambda I) = \lambda^2 - (a+d)\lambda + ad - bc.$$

En ambos casos hay que resolver una ecuación de segundo grado que tiene a lo más dos soluciones reales. Y cuántas son depende del signo de su discriminante, es decir, de lo que queda dentro de la raíz cuadrada al aplicar la fórmula del "chicharronero"; pero en ambos casos queda lo mismo:

$$(d-a)^2 + 4bc = (a+d)^2 - 4(ad-bc). \tag{6.10}$$

Notemos primero que si μ_A invierte orientación, *i.e.*, si $\det A < 0$, entonces, de su última expresión, se tiene que el discriminante es estrictamente positivo y tenemos por tanto dos puntos fijos. Dejaremos el análisis de este caso hasta el final.

Si μ_A preserva orientación, el discriminante (6.10) puede ser negativo (no hay puntos fijos), cero (hay un punto fijo) o positivo (hay dos puntos fijos).

Rotaciones. Como ejemplo de $\mu_A \in \mathbf{Pr}(1)$ sin puntos fijos, podemos tomar a la matriz A de una rotación con ángulo entre 0 y π. Todas las rectas en el plano por 0 se mueven, y $\mathbb{P}^1$, como círculo, simplemente gira. No todas las transformaciones de Moebius sin puntos fijos son tan simples; pero escencialmente sí. Todos sus puntos se mueven en la misma dirección, aunque la "distancia" que se mueven varía.

Translaciones. Si $\mu \in \mathbf{Pr}(1)$ tiene un punto fijo, podemos suponer que es el infinito. Entonces μ es afín, y si no tiene más puntos fijos, es fácil ver que es una traslación (Ejercicio 3.19). Estas transformaciones (con un solo punto fijo) mueven a todos los puntos de $\mathbb{P}^1$ en la misma dirección, excepto al punto fijo, y a los cercanos a él los mueven cada vez más poquito. Si p es el punto fijo, la suceción $\mu^n(q)$ para cualquier $q \neq p$ tiende a p por el mismo lado, y la suceción $\mu^{-n}(q)$ tiende a p por el lado contrario.

Expansiones. La transformación proyectiva $f(x) = kx$ con $k > 1$ (pensando en la recta extendida), fija al 0 y al ∞, y cualquier otro punto se aleja del cero y se acerca a ∞. A 0 se le llama el punto *expulsor* y a ∞ el *atractor;* pues las sucesiones $f^n(x)$ tienden a ∞ y las $f^{-n}(x)$ a 0 para cualquier otro punto x. Toda transformación de Moebius con (exactamente) dos puntos fijos y que preserva orientación es de este tipo. Uno de sus puntos fijos es expulsor y el otro atractor, los demás puntos se mueven de uno hacia el otro.

Consideremos por último a $\mu \in \mathbf{Pr}(1)$ que **invierte orientación**. Ya vimos que entonces tiene dos puntos fijos, p y q digamos. Como $\mu^2 = \mu \circ \mu$ preserva orientación y tiene los mismos dos puntos fijos entonces es una expansión con uno de sus puntos fijos de expulsor, p digamos, y el otro, q, de atractor; diremos entonces que es una **e-reflexión** de p a q. Pero hay otra posibilidad: que $\mu^2 = \text{id}$; y entonces diremos que es **la reflexión** en los puntos p y q. Así, una e-reflexión es una reflexión seguida de una expansión (de ahí el nombre que le hemos dado), y la dirección queda determinada por el valor de $\mu^2(x)$ para cualquier otro punto x (si $\mu^2(x)$ está entre x y q, va de p a q, y si está entre x y p, va de q a p).

Hemos dicho implícitamente que **la reflexión** queda determinada por dos puntos $p \neq q$, sus *puntos espejo,* pero no lo demostramos. Consideremos una transformación de Moebius g que manda a p, q en $0, \infty$ respectivamente y analicemos primero este caso particular. Una $f \in \mathbf{Pr}(1)$ que fije al 0 y al ∞ se escribe $f(x) = kx$ para alguna $k \neq 0$ (es lineal), y si además invierte orientación se debe cumplir que $k < 0$. Como $f^2(x) = k^2 x$, se tiene que $f^2 = \text{id}$ solamente cuando $k = -1$ (para $k < -1$, f es una e-reflexión de 0 a ∞, y para $-1 < k < 0$ es una e-reflexión de ∞ a 0). Sea ρ **la reflexión en** 0 **y** ∞, es decir, $\rho(x) = -x$ (ya hemos demostrado que es única y que ésta es su fórmula). Para obtener la reflexión en p y q, conjugamos con g: sea

$$\rho_{p,q} = g^{-1} \circ \rho \circ g.$$

Se tiene que $\rho_{p,q}(p) = g^{-1}(\rho(g(p))) = g^{-1}(\rho(0)) = g^{-1}(0) = p$, análogamente $\rho_{p,q}(q) = q$; además que $\rho_{p,q}$ invierte orientación pues g y g^{-1} son del mismo tipo y ρ invierte orientación, y por último

$$\rho^2_{p,q} = \left(g^{-1} \circ \rho \circ g\right) \circ \left(g^{-1} \circ \rho \circ g\right) = g^{-1} \circ \rho \circ \rho \circ g = g^{-1} \circ g = \mathrm{id}\,.$$

Lo cuál demuestra que la reflexión en p y q está bien definida y da una fórmula usando el Teorema de 3 en 3. En resumen, se tiene el siguiente lema.

Lema 6.5 *Dados cualesquiera* $p, q \in \mathbb{P}^1$ *con* $p \neq q$*, existe una única* $\rho_{p,q} \in \boldsymbol{Pr}(1)$, *llamada su reflexión, que los tiene como puntos fijos, invierte orientación y cumple que* $\rho^2_{p,q} = \mathrm{id}$. □

A diferencia de la recta euclidiana donde cada par de puntos define un segmento y dos rayos exteriores, dos puntos en la línea proyectiva, que topológicamente es un círculo, la parten en dos segmentos; no hay manera de decir cuál es "el bueno", podemos ir de un punto al otro usándolos indistintamente. Pero la naturaleza es muy sabia, y nos da, en la línea proyectiva, una manera canónica de identificar a estos dos segmentos entre sí, que es **la** reflexión en esos dos puntos. De alguna manera, que se hará explícita cuando veamos la geometría hiperbólica, esta información (una reflexión para cada par de puntos), define a la línea proyectiva pensada a la Klein: un espacio, el círculo, junto con un grupo de transformaciones. Por lo pronto, considera el último de los ejercicios que siguen.

EJERCICIO 6.10 Para cada una de las siguientes matrices, describe qué tipo de transformación de Moebius determina:

$$\textbf{a)}\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \qquad \textbf{b)}\ \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}. \qquad \textbf{c)}\ \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}.$$

$$\textbf{d)}\ \begin{pmatrix} 1 & 0 \\ 3 & -2 \end{pmatrix}. \qquad \textbf{e)}\ \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}.$$

EJERCICIO 6.11 Demuestra que si $f \in \mathbf{Pr}(1)$ es una traslación, entonces existe $g \in \mathbf{Pr}(1)$ tal que $\left(g^{-1} \circ f \circ g\right)(x) = x + 1$.

EJERCICIO 6.12 Demuestra que si $f \in \mathbf{Pr}(1)$ es una expansión entonces existe $g \in \mathbf{Pr}(1)$ tal que $\left(g^{-1} \circ f \circ g\right)(x) = k\,x$ con $k > 1$.

EJERCICIO 6.13 Demuestra que si $f \in \mathbf{Pr}(1)$ es tal que $f^n = \mathrm{id}$ con $n \geq 3$, pero $f^m \neq \mathrm{id}$ para $m < n$, entonces f es una rotación.

EJERCICIO 6.14 Demuestra que la reflexión en 1 y -1 es la función $\nu(x) = 1/x$.

EJERCICIO 6.15 Sea $a \in \mathbb{R} \subset \mathbb{P}^1$; demuestra que la reflexión en a e ∞ es la reflexión (isométrica) en a.

EJERCICIO 6.16 Sean $a, b \in \mathbb{R} \subset \mathbb{P}^1$ dos puntos distintos. Demuestra que la reflexión en a y b tiene la matriz asociada

$$\begin{pmatrix} a+b & -2ab \\ 2 & -a-b \end{pmatrix}.$$

EJERCICIO 6.17 Considera dos parejas distintas, p, q y a, b, de puntos en $\mathbb{P}^1$ (pensado como círculo). Observa que hay tres posibilidades topológicas de colocarlas en el círculo: primero, una pareja está contenida en uno de los segmentos que define la otra (y entonces sucede lo mismo al intercambiarlas); segundo, uno de los puntos de las parejas coincide ($p = a$, digamos), y tercera, las parejas se cruzan, es decir, a y b están en segmentos distintos de p a q. Discute cualitativamente, y demuestra si puedes en algunos casos, qué tipo de transformación es la composición de las reflexiones correspondientes, $\rho_{p,q} \circ \rho_{a,b}$. (Empieza por el segundo caso que es el más simple y es límite, o transición, de los otros dos.)

6.2.6 Armonía

Las reflexiones traen implícita a la extensión natural de armonía que vimos en el Capítulo 2, donde surgió la noción de cuarteta armónica o de par de parejas armónicas en la recta afín; y ahora se completa incluyendo al infinito.

Dada una pareja p, q de puntos en $\mathbb{P}^1$ (piénsese en el ejemplo $1, -1$) decimos que a y b *son armónicos respecto a* p y q, si la reflexión en p y q manda a a en b, es decir, si

$$\rho_{p,q}(a) = b$$

(en el ejemplo, y por un reciente ejercicio, si $b = 1/a$). Obsérvese que "ser armónicos" es realmente una relación de parejas pues $\rho_{p,q} = \rho_{q,p}$ está determinado por la pareja (sin orden), y si $\rho_{p,q}(a) = b$ entonces $\rho_{p,q}(b) = a$ pues $\rho^2_{p,q} = \mathrm{id}$; además, fijando la pareja p, q cualquier otro punto tiene *su armónico*: el complemento $(\mathbb{P} - \{p, q\})$ se aparea en pares armónicos respecto a ellos.

Esta noción de armonía coincide con la euclidiana (y de hecho la extiende), pues vimos que el apareamiento armónico respecto a p, q (reales) consistía en pensar al intervalo p, q como el $-1, 1$ (por una afinidad) e invertir (por la función $x \mapsto 1/x$); sólo el punto medio no tenía par armónico, pero ahora le corresponde el infinito.

Veámos ahora que las transformaciones de Moebius preservan armonía. Sean $f \in \mathbf{Pr}(1)$ y a, b un par armónico respecto a p, q; queremos demostrar que $f(a), f(b)$ es un par armónico respecto a $f(p), f(q)$. Por el mismo argumento del párrafo anterior al Lema 6.5 se tiene que la reflexión en la pareja $f(p), f(q)$ se escribe

$$\rho_{f(p),f(q)} = f \circ \rho_{p,q} \circ f^{-1},$$

y al aplicar esta formula a $f(a)$ se obtiene $f(b)$.

Nos falta ver que la relación de armonía es simétrica, es decir que si $\rho_{p,q}(a) = b$ entonces $\rho_{a,b}(p) = q$. Con una transformación proyectiva podemos mandar p al 0, q al ∞ y a al 1; entonces, por lo anterior, manda b al -1. En este ejemplo se tiene que $0, \infty$ son armónicos respecto a $1, -1$; mandándolos de regreso se tiene entonces lo que queríamos. Podemos resumir en el siguiente teorema.

Teorema 6.6 *Las transformaciones de Moebius preservan cuartetas armónicas y son transitivas en ellas.* □

Por último, notemos que al pensar a la recta proyectiva como líneas en $\mathbb{R}^2$, obtenemos una relación de *armonía en cuartetas de líneas* concurrentes: es que la correspondiente cuarteta de puntos en $\mathbb{P}^1$ sea armónica. Y esto se detecta al intersectarlas con cualquier otra recta que no pasa por el punto de concurrencia, revisando que los puntos de intersección sean armónicos. En particular, si trazamos una paralela a una de ellas, su pareja la debe intersectar en el punto medio de las intersecciones con el otro par (pues hemos mandado a uno de los puntos intencionalmente al infinito, que es el armónico de su punto medio respecto a un segmento).

EJERCICIO 6.18 Demuestra que a, b son armónicos con p, q si y sólo si sus reflexiones conmutan, es decir,

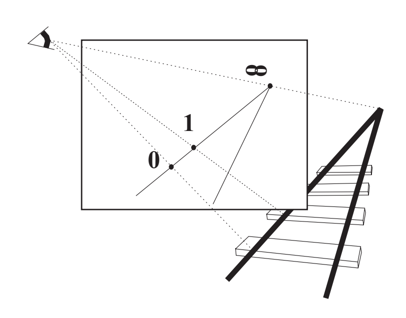
$$\rho_{p,q} \circ \rho_{a,b} = \rho_{a,b} \circ \rho_{p,q}.$$

EJERCICIO 6.19 Demuestra que las bisectrices de cualquier par de rectas es un par armónico a ellas.

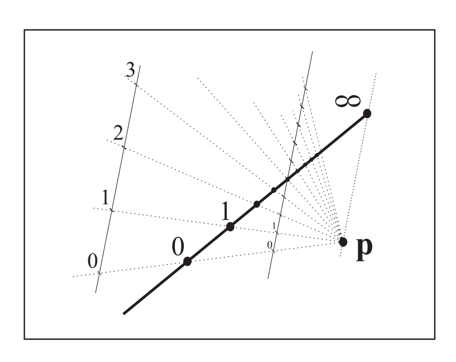
6.3 El problema del pintor II

Recordemos el problema del pintor que quiere pintar en un lienzo una vía de ferrocarril y nos pregunta ¿a cada cuántos centimetros debo dibujar los durmientes? Motivados por este problema definimos la línea proyectiva junto con su grupo de transformaciones y hemos desarrollado una gran teoría con posibilidades de generalizarse a otras dimensiones; pero antes de hacerlo, es justo que hagamos una pausa y le respondamos. Ya podemos darle respuestas numéricas y además constructivas (recetas que usan los arquitectos) basados fundamentalmente en el Teorema de 3 en 3 y en que los elementos del grupo de transformaciones de Moebius se pueden ver como proyecciones. Pero además, para desarrolllar la intuición y como motivación para entrarle al plano proyectivo, veremos que nuestros métodos funcionan al ampliar el problema a dibujar una cuadrícula en un piso.

El riel (izquierdo, para fijar ideas) se dibuja como una línea recta en el lienzo, y esta línea tiene un punto distinguido: el punto de fuga en donde los dos rieles de la vía parecen juntarse. Con ese punto y cualquier otro, el del durmiente 0 digamos, se define esa recta. Sabemos que la función que nos dice a dónde debe ir el durmiente n es una transformación de Moebius, y que ésta estará determinada por la imagen del durmiente 0, el punto de fuga, ∞, y algún otro punto. Así que si el pintor nos da el ∞, el durmiente 0 y el durmiente 1 en el lienzo, podríamos calcular los coeficientes de la función de Moebius (después de medir cuidadosamente las distancias en el lienzo), programarla en una calculadora y regresársela al pintor para que continúe con su trabajo. La calculadora le dirá exactamente dónde debe dibujar al durmiente 2, al 3, etc. Problema resuelto.



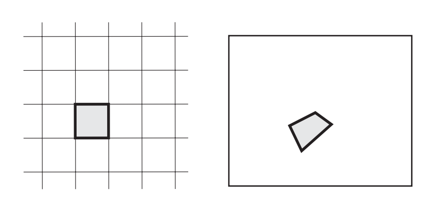
Pero no para el pintor, que requeriría de un matemático de cabecera. Él quisiera una solución dentro de sus posibilidades y más cercana a las técnicas que maneja. Veamos de nuevo. También hemos demostrado que cualquier transformación de Moebius se puede ver como la proyección de una línea en otra. Queremos proyectar una escala afín (la del ferrocarril real con unidad de durmientes) a la escala proyectiva del lienzo desde el punto fijo de visión del pintor. Puesto que es una transformación de Moebius entre rectas, entonces debería ser posible diseñarle una construcción geométrica para que con líneas y puntos auxiliares en el lienzo reconstruya la escala proyectiva (del riel en el lienzo) a partir de una escala afín (la de su regla T, digamos) sin necesidad de una calculadora, o de estarle atinando a ojo.



Suponemos entonces que tenemos definido el riel en el lienzo junto con sus tres puntos distinguidos. Fijemos arbitrariamente un punto de proyección $\mathbf{p}$ (fuera del riel). La recta ℓ que pasa por $\mathbf{p}$ e ∞ determina la dirección de la recta cuya escala afín vamos a proyectar (pues el infinito de ésta debe proyectarse en el infinito que tenemos dado). Para cualquier recta paralela a ℓ, sus puntos de intersección con las rectas por $\mathbf{p}$ y el 0 y por $\mathbf{p}$ y el 1 (nos referimos al 0 y al 1 en el riel del lienzo) determinan la escala. Con un compás se puede reproducir esta escala afín que, proyectándola desde $\mathbf{p}$, nos da la escala proyectiva en el riel. Lo maravilloso de la teoría que hemos desarrollado es que demuestra que este procedimiento no depende del punto de proyección o de la recta a proyectar que escojamos; además, demuestra también que coincide con el primer método (el de la calculadora), y, por si fuera poco, que estos dos métodos coinciden con el original de proyectar al riel verdadero sobre el riel del lienzo.

Con este método, el pintor puede sacar la escala proyectiva del riel derecho (a partir del infinito, que comparten los rieles y los puntos donde se ven el durmiente 0 y el durmiente 1), luego la del riel izquierdo, y, detallando durmientes, concluir con una

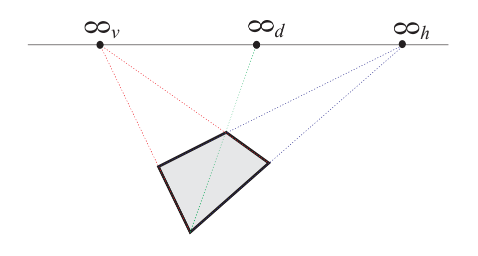
buena perspectiva de la vía de tren. Pero aún puede simplificarse el procedimiento si la vía del tren está lo suficientemente ladeada para que los durmientes también estén en perspectiva. En este caso, que analizaremos a continuación, conviene plantearse un problema que parece más difícil: el de dibujar un piso cuadriculado en perspectiva.



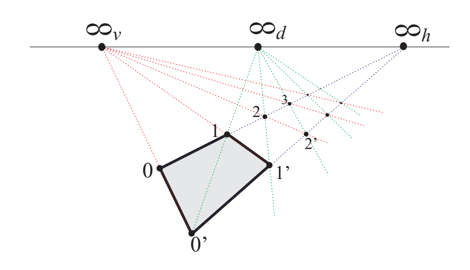
Queremos ahora proyectar un piso cuadriculado sobre un lienzo. Si le pedimos al pintor que dibuje, a ojo de buen pintor, una simple loseta, podremos reconstruir todo el piso.

Pensemos que la franja del piso que contiene a la loseta y que se aleja a la derecha (*horizontal*, llamémosla) es una vía de ferrocarril con durmientes verticales. Las dos líneas horizontales parecen converger a un punto de fuga que en el lienzo es un punto bien definido (el infinito horizontal, ∞_h) como el punto de intersección de las dos rectas definidas por los lados opuestos de la loseta ya dibujada.

Por la misma razón, las líneas de la franja vertical convergen a otro punto (el infinito vertical, ∞_v). Si estos dos puntos de fuga caen en el lienzo (que es el caso que analizaremos), determinan una línea llamada horizonte (definida en el lienzo como la intersección de éste con el plano paralelo al piso que pasa por el ojo del pintor, en general, una línea horizontal en el lienzo justo a la misma altura del piso que el ojo de proyección).



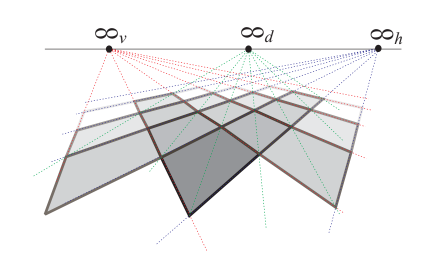
Sabemos que todas las líneas horizontales del piso se dibujan como líneas en el lienzo que confluyen al punto ∞_h; no importa el ancho de la vía del tren. Y por la misma razón, las verticales confluyen en ∞_v. Pero las horizontales y las verticales no gozan de ningún beneficio (excepto el de definir la cuadrícula); cualquier otro haz de rectas paralelas se proyecta en un haz que confluye en algún punto del horizonte. Así que podemos tomar una diagonal de la loseta dada, que intersecta al horizonte en un punto que llamaremos infinito diagonal, ∞_d, que está entre ∞_v e ∞_h.



Si llamamos 0 y 1, como en el caso del ferrocarril, a los dos vértices en un lado horizontal de la loseta, y $0'$ y $1'$ a los correspondientes del lado opuesto, de tal manera que la diagonal pase por $0'$ y 1, entonces la línea que va de $1'$ a ∞_d es justo donde se dibuja la diagonal de la siguiente loseta; cuya intersección con la horizontal por 0 y 1 nos da el vértice 2 correspondiente a la siguiente loseta. Trazando ahora la vertical por 2, su intersección con el otro riel nos da el vertice $2'$, completando así la información para dibujar la loseta 1 (empezamos con la 0). Es claro que este proceso se puede iterar para encontrar los durmientes que queramos del ferrocarril horizontal.

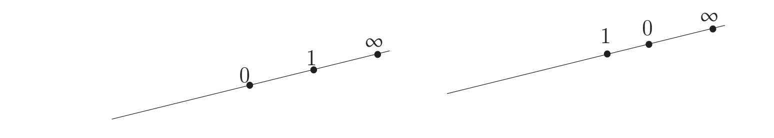
Pero de manera análoga, se puede completar el ferrocarril vertical y por ahí enmedio van apareciendo los vértices de más losetas. E inclusive de losetas "negativas".

Este método se conoce desde el Renacimiento. Obsérvese que ya ni siquiera necesitamos compás, con pura regla funciona. Y funciona bien, lo cual indica que debe haber teoremas que lo sustenten. Por ejemplo ¿porqué es que los puntos de intersección de las diagonales con las verticales siguientes ($0'$ y 1 con $2'$ y 2, $1'$ y 2 con $3'$ y 3, $2'$ y 3 con $4'$ y 4, etc.) quedan todos alineados con ∞_h? Desde el Renacimiento se sabe por experiencia que así resulta, pero los fundamentos matemáticos tardaron siglos en desarrollarse. Por el momento dejemos esto como motivación para estudiar y definir al plano proyectivo, dejando también al pintor contento con los métodos que sí puede aplicar.



EJERCICIO 6.20 En el primer método geométrico para encontrar la escala en el riel, el punto de proyección $\mathbf{p}$ se fijó arbitrariamente. Supón que la que se fija arbitrariamente es la recta ℓ (digamos que es la orilla del restirador donde trabaja el pintor). ¿Cómo encuentras el punto de proyección y la escala afín en ℓ para proyectar?

EJERCICIO 6.21 Con regla y compás completa las siguientes escalas proyectivas añadiendo los puntos que corresponden a $-1, 2$ y 3. Haz otras.



6.4 El plano proyectivo

Nuestra definición del plano proyectivo será analítica, generalizando la idea de coordenadas homogéneas que acabamos de desarrollar. Esto nos permitirá hacer cálculos y definir objetos geométricos de manera precisa, pero peca de abstracta. En particular nos interesa en esta sección establecer la noción de línea y las propiedades básicas de incidencia. Para tener una intuición geométrica del plano proyectivo, que es muy importante, veremos en la siguiente sección otros modelos más geométricos. Es preciso remarcar, una y otra vez, que cada modelo geométrico tiene ventajas y desventajas, no hay una manera de "ver" todo el plano proyectivo de golpe, como creemos tener la idea del plano euclidiano. Pero observemos que esto último tampoco es cierto. Lo que intuitivamente sabemos del plano euclidiano, es por pequeñas porciones de él, el pizarrón, el piso o una hoja de papel, y de ahí extrapolamos diciendo "todo es así". Lo mismo sucede con el plano proyectivo, en pequeñas porciones es como el euclidiano pero también incluye cierto comportamiento al infinito que se convierte en parte integral de él.

6.4.1 Coordenadas homogéneas

Dado $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$ tal que $\mathbf{x} \neq \mathbf{0}$ (esto es, tal que alguna de las tres coordenadas de $\mathbf{x}$ es distinta de cero), sea

$$[\mathbf{x}] = [x : y : z] := \{(tx, ty, tz) \mid t \in \mathbb{R},\ t \neq 0\}.$$

Es decir, $[\mathbf{x}]$ es la recta por el origen con dirección $\mathbf{x}$, menos el origen. Se tiene ....Dibujo entonces que

$$[\mathbf{x}] = [t\,\mathbf{x}] \text{ para cualquier } t \neq 0,$$

y ésta es la propiedad clave. Además consideraremos a $[\mathbf{0}]$ o $[0 : 0 : 0]$ como una expresión sin sentido.

Definimos al *plano proyectivo* $\mathbb{P}^2$ como

$$\mathbb{P}^2 := \left\{[\mathbf{x}] \mid \mathbf{x} \in \mathbb{R}^3, \mathbf{x} \neq \mathbf{0}\right\}.$$

Nos referimos a los elementos de $\mathbb{P}^2$ como *puntos* y a la expresión $[x : y : z]$ o $[\mathbf{x}]$ como un punto de $\mathbb{P}^2$ dado por *coordenadas homogéneas.*

Nótese que las coordenadas homogéneas de un punto en $\mathbb{P}^2$ no son únicas (a diferencia de $\mathbb{R}^2$, donde sí lo son). Hay cierta ambigüedad, pero no es tan grave, pues

$$[\mathbf{x}] = [\mathbf{y}] \quad \text{si y sólo si } \mathbf{y} = t\,\mathbf{x} \text{ para alguna } t \neq 0;$$

así que por ejemplo, $[1 : 2 : 3] = [2 : 4 : 6]$, o bien $[1 : 0 : 0] = [8 : 0 : 0] = \left[2^8 : 0 : 0\right]$. Como en el caso de $\mathbb{R}^2$ (o de $\mathbb{R}^n$) los puntos de $\mathbb{P}^2$ tienen coordenadas, y esto nos permitirá hacer geometría con métodos analíticos, pero habrá que tener cuidado de que las definiciones no dependan de coordenadas específicas sino de las homogéneas.

EJERCICIO 6.22 Generaliza nuestra definición de $\mathbb{P}^1$ y de $\mathbb{P}^2$ para definir $\mathbb{P}^3$ y $\mathbb{P}^n$.

6.4.2 Rectas proyectivas

Tenemos que *los puntos de $\mathbb{P}^2$ corresponden a las rectas en $\mathbb{R}^3$ que pasan por el origen* —si $\mathbf{x} \in \mathbb{R}^3 - \{\mathbf{0}\}$, se puede pensar a $[\mathbf{x}]$ como punto en $\mathbb{P}^2$ ($[\mathbf{x}] \in \mathbb{P}^2$) o bien como su correspondiente línea por el origen en $\mathbb{R}^3$ ($[\mathbf{x}] \subset \mathbb{R}^3$). Podemos definir entonces a los subconjuntos distinguidos de $\mathbb{P}^2$ que son sus *rectas o líneas* correspondiendo a los planos por el origen en $\mathbb{R}^3$; es decir, una *recta* en $\mathbb{P}^2$ es el conjunto de puntos cuyas correspondientes rectas en $\mathbb{R}^3$ forman un plano. Tiene sentido, pues en la Sección 6.2.4 vimos que los puntos de una recta proyectiva corresponden biunívocamente a las rectas por el origen en un plano. Puesto que los planos en $\mathbb{R}^3$ están dados por una ecuación lineal, podemos ser más explícitos:

Sea $\mathbf{n}$ un vector no nulo en $\mathbb{R}^3$; definimos la *recta dada por* $\mathbf{n}$ en $\mathbb{P}^2$ como

$$\ell_{\mathbf{n}} := \left\{[\mathbf{x}] \in \mathbb{P}^2 \mid \mathbf{n} \cdot \mathbf{x} = 0\right\}. \tag{6.11}$$

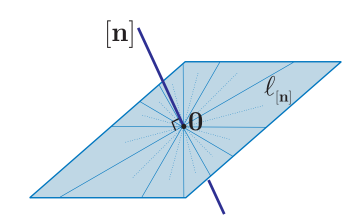
Está bien definida pues si tomamos $t \neq 0$, entonces

$$\mathbf{n} \cdot \mathbf{x} = 0 \;\Leftrightarrow\; \mathbf{n} \cdot t\mathbf{x} = 0,$$

pues $\mathbf{n} \cdot t\mathbf{x} = t(\mathbf{n} \cdot \mathbf{x})$, lo cual implica que no importa qué representante de un punto tomemos ($[\mathbf{x}]$ o $[t\mathbf{x}]$) la condición para pertenecer al conjunto da el mismo resultado (es cierta o no dependiendo del punto en $\mathbb{P}^2$, no del representante en $\mathbb{R}^3$). Si $\mathbf{n} = (a, b, c)$ entonces $\ell_\mathbf{n}$ se describe por coordenadas homogéneas como

$$\ell_\mathbf{n} := \left\{[x : y : z] \in \mathbb{P}^2 \mid ax + by + cz = 0\right\}.$$

Y podemos decir que las rectas de $\mathbb{P}^2$ son precisamente los conjuntos $\ell_\mathbf{n}$ para $\mathbf{n} \in \mathbb{R}^3$, $\mathbf{n} \neq \mathbf{0}$. Notemos además que si $t \neq 0$, entonces $\ell_\mathbf{n} = \ell_{t\mathbf{n}}$, (pues $t\mathbf{n} \cdot \mathbf{x} = t(\mathbf{n} \cdot \mathbf{x}) = 0 \Leftrightarrow \mathbf{n} \cdot \mathbf{x} = 0$). Así que la recta proyectiva $\ell_\mathbf{n}$ depende no de $\mathbf{n}$ (un vector no nulo en $\mathbb{R}^3$) sino de $[\mathbf{n}]$ (el correspondiente punto en $\mathbb{P}^2$); entonces podemos definir para $[\mathbf{n}] \in \mathbb{P}^2$ su *recta dual*



$$\ell_{[\mathbf{n}]} := \ell_\mathbf{n}.$$

Dicho de otra manera, a cada recta por el origen en $\mathbb{R}^3$ ($[\mathbf{n}]$) le asociamos su plano perpendicular que pasa por el origen ($\ell_{[\mathbf{n}]}$). Pero además, como cada plano por el origen tiene una única recta perpendicular por el origen, se tiene que

$$\ell_\mathbf{n} = \ell_\mathbf{m} \quad \Leftrightarrow \quad [\mathbf{n}] = [\mathbf{m}].$$

Así que hemos establecido una biyección entre las rectas de $\mathbb{P}^2$ y los puntos de $\mathbb{P}^2$

$$\ell_{[\mathbf{n}]} \longleftrightarrow [\mathbf{n}]$$

que llamaremos *dualidad*. La dualidad cumple propiedades muy bonitas que vienen de una básica que dice que un punto está en una recta si y sólo si su recta dual contiene al punto dual; es decir, dados $[\mathbf{p}]$ y $[\mathbf{q}]$ en $\mathbb{P}^2$,

$$[\mathbf{p}] \in \ell_{[\mathbf{q}]} \;\Leftrightarrow\; [\mathbf{q}] \in \ell_{[\mathbf{p}]}, \tag{6.12}$$

cuya demostración es obvia pues ambas condiciones son, por definición, equivalentes a $\mathbf{p} \cdot \mathbf{q} = 0$.

EJERCICIO 6.23 ¿Cómo definirías planos en $\mathbb{P}^3$?

EJERCICIO 6.24 ¿Cómo definirías rectas en $\mathbb{P}^3$?

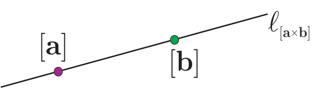
EJERCICIO 6.25 ¿Habrá dualidad en $\mathbb{P}^3$? ¿Entre qué y qué?

6.4.3 Axiomas de incidencia y el quinto postulado

Demostremos ahora las propiedades más elementales del plano proyectivo, que son las de *incidencia*, y se refieren a las relaciones "estar en" o "pasar por" que hay entre puntos y rectas.

Lema 6.7 *Por cualesquiera dos puntos de* $\mathbb{P}^2$ *pasa una recta, y si son distintos ésta es única.*

Demostración. Sean $[\mathbf{a}]$ y $[\mathbf{b}]$ puntos en $\mathbb{P}^2$. Si $[\mathbf{a}] = [\mathbf{b}]$ entonces para cualquier $[\mathbf{n}] \in \ell_{[\mathbf{a}]}$ se tiene que $[\mathbf{a}] \in \ell_{[\mathbf{n}]}$ por (6.12). Si $[\mathbf{a}] \neq [\mathbf{b}]$, sea $\mathbf{n} = \mathbf{a} \times \mathbf{b}$, entonces $\mathbf{n} \neq \mathbf{0}$ y se tiene que $[\mathbf{a}], [\mathbf{b}] \in \ell_{[\mathbf{n}]}$, pues $\mathbf{n} \cdot \mathbf{a} = \mathbf{n} \cdot \mathbf{b} = 0$. Además, esta recta es única pues $\mathbf{n}$ genera a la recta perpendicular al plano que generan los vectores $\mathbf{a}$ y $\mathbf{b}$. □



Y tenemos el resultado dual en el siguiente lema.

Lema 6.8 *Dos rectas distintas de* $\mathbb{P}^2$ *tienen un único punto en común.*

Demostración. Dadas $\ell_{[\mathbf{a}]}$ y $\ell_{[\mathbf{b}]}$ rectas en $\mathbb{P}^2$, si $\ell_{[\mathbf{a}]} = \ell_{[\mathbf{b}]}$ no hay nada que demostrar. Si $\ell_{[\mathbf{a}]} \neq \ell_{[\mathbf{b}]}$ (es decir, si $[\mathbf{a}] \neq [\mathbf{b}]$), sea $\mathbf{n} = \mathbf{a} \times \mathbf{b}$, entonces $\mathbf{n} \neq \mathbf{0}$, pues $\mathbf{a}$ y $\mathbf{b}$ son linealmente independientes; y se tiene que $[\mathbf{n}] \in \ell_{[\mathbf{a}]} \cap \ell_{[\mathbf{b}]}$, pues $\mathbf{a} \cdot \mathbf{n} = \mathbf{b} \cdot \mathbf{n} = 0$. □

Nótese que este lema consiste simplemente en pasar a lenguaje de puntos en $\mathbb{P}^2$ el hecho de que dos planos por el origen en $\mathbb{R}^3$ se intersectan en una línea; y el anterior que dos líneas por el origen generan un plano.

Notación. A veces será más fácil referirnos a los puntos de $\mathbb{P}^2$ por una sola letra como $\mathbf{p}$ o $\mathbf{q}$; en este caso, sabemos que podemos recurrir a sus coordenadas homogéneas o a algún vector representante, pero en el contexto éstos no son relevantes. También será útil tener una notación para *la recta que generan dos puntos*:

$$\langle \mathbf{p}, \mathbf{q} \rangle$$

denotará la recta que pasa por $\mathbf{p}$ y $\mathbf{q}$. Así que hemos demostrado que si $[\mathbf{a}] \neq [\mathbf{b}]$ entonces

$$\langle [\mathbf{a}], [\mathbf{b}] \rangle = \ell_{[\mathbf{a}\times\mathbf{b}]}.$$

Obsérvese también que las dos demostraciones anteriores son casi la misma, y esto no es casualidad. Se debe al *principio de dualidad* (6.12) que nos permite traducir cualquier resultado a su versión dual intercambiando "recta" por "punto" y la relación "pasar por" por la de "estar en". Pero de esto hablaremos más adelante con ejemplos más interesantes a la mano. Por el momento queremos insistir en la diferencia básica con el plano euclidiano.

En el plano euclidiano se cumple el Lema 6.7, pero no el 6.8. De hecho este último contradice al Quinto Postulado; establece que en el plano proyectivo no hay paralelismo, que cualquier par de rectas se comporta de la misma manera: se intersectan en un único punto.

EJERCICIO 6.26 Sean $[\mathbf{a}] = [1:0:-1]$, $[\mathbf{b}] = [2:1:-1]$, $[\mathbf{c}] = [0:1:-1]$ y $[\mathbf{d}] = [1:-1:3]$. Encuentra el punto de intersección de las rectas que pasan por $[\mathbf{a}]$ y $[\mathbf{b}]$ y por $[\mathbf{c}]$ y $[\mathbf{d}]$; es decir

$$\langle [\mathbf{a}], [\mathbf{b}] \rangle \cap \langle [\mathbf{c}], [\mathbf{d}] \rangle .$$

EJERCICIO 6.27 Los cuatro puntos del ejercicio anterior generan seis rectas cuyas intersecciones dan siete puntos. ¿Cuáles son los dos que no hemos dado explícitamente?

EJERCICIO 6.28 En el plano $z = 1$ de $\mathbb{R}^3$ puedes encontrar representantes de los puntos del Ejercicio 6.26; haz un dibujo en él (como papel con coordenadas x, y) de los objetos mencionados en los ejercicios anteriores.

6.4.4 Parametrización de rectas proyectivas

En el plano euclidiano, toda recta es copia de la recta real mediante una parametrización. Veremos ahora el análogo: que en el plano proyectivo toda recta es una copia de la recta proyectiva.

Dados los puntos $[\mathbf{a}], [\mathbf{b}] \in \mathbb{P}^2$, tales que $[\mathbf{a}] \neq [\mathbf{b}]$, tenemos una *parametrización* de la recta que pasa por ellos dada por la función

$$\begin{array}{rcl} \mathbb{P}^1 & \to & \mathbb{P}^2 \\ [x:y] & \mapsto & [x\mathbf{a} + y\mathbf{b}] \,; \end{array}$$

...Dibujo

que corresponde a la manera natural de generar un plano por dos vectores dados. Esta función es inyectiva y nos dice que cada recta del plano es como la recta proyectiva modelo, o abstracta, $\mathbb{P}^1$. Sin embargo no está unívocamente determinada por los puntos $[\mathbf{a}]$ y $[\mathbf{b}]$ (aunque sí por sus representantes $\mathbf{a}$ y $\mathbf{b}$). Si tomamos $t\,\mathbf{a}$ en vez de $\mathbf{a}$, con $t \neq 0$ por supuesto, entonces la asignación $[x:y] \mapsto [x\,(t\,\mathbf{a}) + y\mathbf{b}]$ sigue mandando al $[1:0]$ en $[\mathbf{a}]$ y al $[0:1]$ en $[\mathbf{b}]$, pero en el resto de los puntos es diferente (si $t \neq 1$). Esto corresponde a que una transformación proyectiva de $\mathbb{P}^1$ está determinada por su valor en tres puntos, no sólo en dos. Así que dado cualquier otro

punto $[\mathbf{c}] \in \langle[\mathbf{a}], [\mathbf{b}]\rangle$ se tiene que $\mathbf{c} = t\,\mathbf{a} + s\,\mathbf{b}$ para algunos t y s que son distintos de cero si $[\mathbf{a}] \neq [\mathbf{c}] \neq [\mathbf{b}]$. Y entonces la función

$$\begin{array}{rcl} \mathbb{P}^1 & \to & \mathbb{P}^2 \\ [x:y] & \mapsto & [x(t\,\mathbf{a}) + y\,(s\,\mathbf{b})], \end{array}$$

que manda a la terna canónica $[1:0], [0:1], [1:1]$ en la terna $[\mathbf{a}], [\mathbf{b}], [\mathbf{c}]$ sí está unívocamente determinada.

Estas inclusiones de $\mathbb{P}^1$ en $\mathbb{P}^2$ son ejemplos de *funciones proyectivas*. Tenemos entonces que todas las rectas de $\mathbb{P}^2$ son como nuestro modelo ideal de recta proyectiva $\mathbb{P}^1$.

EJERCICIO 6.29 Sean $[\mathbf{a}] = [1:0:-1], [\mathbf{b}] = [2:1:-1]$ y $[\mathbf{c}] = [0:-1:0]$. Demuestra que son colineales y encuentra explícitamente la función proyectiva de $\mathbb{P}^1$ en $\mathbb{P}^2$ que manda al $[1:0]$ en $[\mathbf{a}]$, al $[0:1]$ en $[\mathbf{b}]$ y al $[1:1]$ en $[\mathbf{c}]$.

EJERCICIO 6.30 ¿Puedes definir líneas en $\mathbb{P}^3$? ¿Funciones proyectivas de $\mathbb{P}^2$ en $\mathbb{P}^3$?

6.5 Modelos del plano proyectivo

6.5.1 Completación de $\mathbb{R}^2$

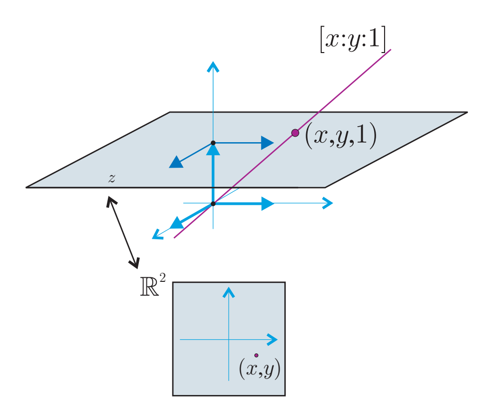
Consideremos al plano en $\mathbb{R}^3$ dado por la ecuación $z = 1$; llamémosle Π_z. Tiene una parametrización natural tomando como origen al tercer vector canónico $\mathbf{e}_3$ y como generadores a $\mathbf{e}_1$ y $\mathbf{e}_2$; es decir, lo podemos identificar con $\mathbb{R}^2$ mediante la inclusión:

$$\begin{array}{rcl} \mathbb{R}^2 & \to & \Pi_z \subset \mathbb{R}^3 \\ (x,y) & \mapsto & (x,y,1). \end{array}$$

Consideremos también al siguiente conjunto de $\mathbb{P}^2$:

$$\mathbb{A}_z := \left\{[x:y:z] \in \mathbb{P}^2 \mid z \neq 0\right\}.$$

Consiste de todos los puntos de $\mathbb{P}^2$ que no cumplen $z = 0$; es decir, es el complemento en el plano proyectivo de la recta $\ell_{[\mathbf{e}_3]}$; esto es, $\mathbb{A}_z = \mathbb{P}^2 - \ell_{[\mathbf{e}_3]}$. Entonces, tenemos una biyección natural



$$\begin{array}{rcl} \Pi_z & \to & \mathbb{A}_z \\ (x,y,1) & \mapsto & [x:y:1], \end{array}$$

cuya inversa, que está bien definida, es

$$\begin{array}{rcl} \mathbb{A}_z & \to & \Pi_z \\ [x:y:z] & \mapsto & (x/z, y/z, 1). \end{array}$$

Si pensamos a los puntos de $\mathbb{P}^2$, geométricamente, como rectas por el origen en $\mathbb{R}^3$, entonces $\mathbb{A}_z$ consiste de todas estas rectas excepto las horizontales (en el plano $z = 0$, que corresponden a puntos de $\ell_{[\mathbf{e}_3]}$). Cada una de las rectas $[\mathbf{x}] \in \mathbb{A}_z$ intersecta a Π_z en un único punto. Y ésta es la biyección $\mathbb{A}_z \to \Pi_z$, $([\mathbf{x}] \mapsto [\mathbf{x}] \cap \Pi_z)$.

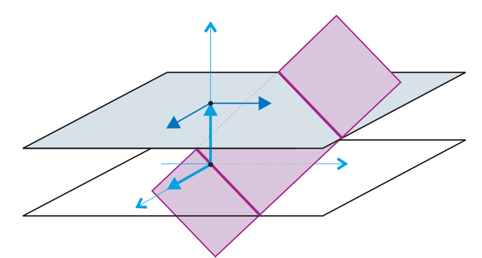
Componiendo las dos biyecciones que acabamos de definir, podemos identificar a $\mathbb{R}^2$ con el complemento de una recta en $\mathbb{P}^2$:

$$\begin{aligned} \mathbb{R}^2 &\to \mathbb{A}_z = \mathbb{P}^2 - \ell_{[\mathbf{e}_3]} \\ (x, y) &\mapsto [x : y : 1] \,. \end{aligned} \tag{6.13}$$

A esta inclusión de $\mathbb{R}^2$ en el plano proyectivo se le llama una *carta coordenada* (pues da coordenadas locales para un pedazo de $\mathbb{P}^2$), y a ésta en particular la llamaremos la *carta coordenada canónica*. Observemos además que las rectas por el origen en $\mathbb{R}^3$ cercanas a las horizontales intersectan en puntos muy lejanos al plano Π_z y, por tanto, corresponden a puntos de norma muy grande en $\mathbb{R}^2$. De tal manera que si a $\mathbb{R}^2$ le "pegamos" una recta proyectiva al infinito (llamémosla $\ell_\infty = \ell_{[\mathbf{e}_3]}$) obtenemos $\mathbb{P}^2$,

$$\mathbb{P}^2 = \mathbb{R}^2 \cup \ell_\infty \,.$$

Veamos ahora que las rectas de $\mathbb{R}^2$ y de $\mathbb{P}^2$ se corresponden. Una recta ℓ en $\mathbb{R}^2$ está dada por una ecuación $ax + by = c$:



$$\ell = \left\{ (x, y) \in \mathbb{R}^2 \mid ax + by = c \right\} .$$

Por (6.13) sus puntos van a los puntos de la recta dual a $[a : b : -c]$:

$$\left\{ [x : y : z] \in \mathbb{P}^2 \mid ax + by - cz = 0 \right\} ,$$

y al único punto que no le "pegan" es a $[-b : a : 0]$; que es justo la intersección de esta última recta con la recta al infinito.

Geométricamente, una recta en $\mathbb{R}^2$ pensado como Π_z genera junto con el origen (de $\mathbb{R}^3$) un plano que corresponde a una recta en $\mathbb{P}^2$; lo único que no se "ve" en Π_z es la intersección de ese plano con el plano horizontal ($z = 0$), que es $[-b : a : 0]$. Obsérvese además, que **las rectas paralelas en** $\mathbb{R}^2$ (obtenidas al variar c en la descripción anterior) **pasan por el infinito en este mismo punto**.

...Dibujo

Inversamente, una recta en $\mathbb{P}^2$ corresponde a un plano por el origen en $\mathbb{R}^3$, y todos ellos, excepto el horizontal ($z = 0$), intersectan a Π_z en una recta.

Podemos resumir diciendo que el plano proyectivo se obtiene del euclidiano añadiendo un *punto al infinito* por cada clase de paralelismo de rectas. Las rectas del proyectivo son las rectas del euclidiano a las que se adjunta el punto al infinito de su clase de paralelismo (convirtiéndolas en rectas proyectivas), y además, todos los puntos al infinito constituyen una nueva recta, llamada "la recta al infinito".

6.5.2 Cartas afines

En la sección anterior demostramos que $\mathbb{P}^2$ se puede ver como $\mathbb{R}^2$ junto con una recta al infinito ℓ_∞. Podría parecer que esta recta tiene algo especial, pero no es asi. Escogimos la tercera coordenada para definir el plano $\Pi_z \subset \mathbb{R}^3$ y el subconjunto $\mathbb{A}_z$ de $\mathbb{P}^2$ por conveniencia. Podemos usar la segunda coordenada y para definir $\Pi_y : y = 1$; $\mathbb{A}_y := \left\{[x : y : z] \in \mathbb{P}^2 \mid y \neq 0\right\}$, de tal manera que $\mathbb{A}_y = \mathbb{P}^2 - \ell_{[\mathbf{e}_2]}$ y tenemos una nueva carta coordenada

$$\begin{array}{rcl} \mathbb{R}^2 & \to & \mathbb{A}_y \\ (x, z) & \mapsto & [x : 1 : z]\,, \end{array}$$

cuya inversa es

$$\begin{array}{rcl} \mathbb{A}_y & \to & \mathbb{R}^2 \\ [x : y : z] & \mapsto & (x/y, z/y). \end{array}$$

En esta nueva carta coordenada, lo que antes era la recta al infinito se ve como el primer eje coordenado de $\mathbb{R}^2$, es decir, corresponde a la recta $\left\{(x, z) \in \mathbb{R}^2 \mid z = 0\right\}$; el origen de $\mathbb{R}^2$ corresponde al punto $[0 : 1 : 0] = [\mathbf{e}_2]$, y lo que ahora no se ve (la nueva recta al infinito) es su dual $\ell_{[\mathbf{e}_2]}$. Si pensamos a estas cartas coordenadas como fotografías de porciones grandes de $\mathbb{P}^2$, con estas dos cartas hemos retratado a todos los puntos, con excepción de uno: $[1 : 0 : 0] = [\mathbf{e}_1]$, que en ambas cartas quedó en el infinito. Pero debe ser claro ahora que si hacemos lo mismo con la primera coordenada x, obtenemos una nueva carta $\mathbb{R}^2 \to \mathbb{A}_x = \mathbb{P}^2 - \ell_{[\mathbf{e}_1]}$ con coordenadas naturales (y, z), donde $[\mathbf{e}_1]$ aparece justo en el origen. ...Dibujo

El nombre de "carta" que hemos estado usando impunemente, ahora sí cobra sentido: viene de la palabra antigua para referirse a "mapa". Y con las tres cartas que hemos definido armamos un *atlas* que describe a todo el plano proyectivo. Nos dice que **el plano proyectivo es localmente como el euclidiano** y las rectas de ambos se corresponden (también localmente).

Hemos usado el apellido "coordenada" para las cartas, pues son funciones explícitas de $\mathbb{R}^2$ en $\mathbb{P}^2$ que nos dan coordenadas locales al quitar una de las tres rectas $\ell_{[\mathbf{e}_i]}$. Pero, además de su sencillez analítica, no hay nada de especial en estas tres rectas. La idea funciona para que cualquier recta se vaya al infinito, pero entonces lo que nos queda no tiene coordenadas naturales (para hacerse corresponder a $\mathbb{R}^2$) pues no habrá origen ni ejes canónicos. Veamos. ...Dibujo

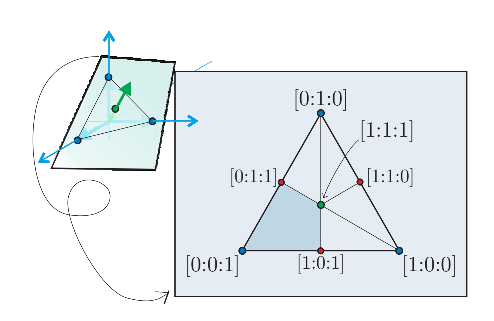
Dada una recta $\ell_{[\mathbf{n}]} \subset \mathbb{P}^2$ consideremos el plano $\Pi_{\mathbf{n}} : \mathbf{n}\cdot\mathbf{x} = 1$ (el 1 podría cambiarse por cualquier constante distinta de cero, pero así será más fácil pues cambiar al 1 corresponde a alargar $\mathbf{n}$). Cualquier recta por el origen $[\mathbf{x}] \in \mathbb{P}^2$ que cumpla que $[\mathbf{x}] \notin \ell_{[\mathbf{n}]}$ (es decir, que $\mathbf{n} \cdot \mathbf{x} \neq 0$) intersecta a $\Pi_{\mathbf{n}}$ justo en un punto. A saber, en $(\mathbf{n} \cdot \mathbf{x})^{-1}\mathbf{x}$. De tal manera que $\Pi_{\mathbf{n}}$ puede identificarse con $\mathbb{P}^2 - \ell_{[\mathbf{n}]}$ por medio de la

función

$$\begin{array}{rcl} \Pi_n & \to & \mathbb{P}^2 - \ell_{[n]} \\ x & \mapsto & [x] \, . \end{array}$$

Puesto que Π_n es un plano afín (no tiene coordenadas naturales), a ésta la llamamos *carta afín*. Lo cual implica que cualquier recta de $\mathbb{P}^2$ funciona como recta al infinito en la carta afín que abarca a su complemento.

Como ejemplo importante, veamos al plano que contiene a los tres vectores canónicos. Sea $\mathbf{u} = (1, 1, 1)$, y sea $\Pi : \mathbf{u} \cdot \mathbf{x} = 1$. Es decir, Π está dado por la ecuación $x+y+z = 1$. De tal manera que $\mathbf{e}_i \in \Pi$ para $i = 1, 2, 3$. Además forman un triángulo equilátero con $(1/3)\mathbf{u}$ en el baricentro. Este plano contiene a un representante de cada $[x : y : z] \in \mathbb{P}^2$ que cumpla $x + y + z \neq 0$; a saber, $(x/(x+y+z), y/(x+y+z), z/(x+y+z))$. Esta carta afín es simétrica respecto a las tres coordenadas; las rectas al infinito de las tres cartas coordenadas que vimos al principio son ahora las rectas del triángulo básico. Nos será útil más adelante; entonces nos referiremos a ella como *la carta afín canónica*. En ella también se ve la "loseta canónica" (los puntos $(0, 0)$, $(1, 0)$, $(0, 1)$, $(1, 1)$ de la carta coordenada canónica) y los siete puntos en que se basó el trazado de pisos del pintor. Nótese que al poner coordenadas proyectivas ya no nos preocupamos de que las coordenadas sumen 1, pero que a partir de ellas, las coordenadas baricéntricas correspondientes se obtienen fácilmente.



EJERCICIO 6.31 Describe la intersección de dos cartas afines.

EJERCICIO 6.32 Sea $\varphi_3 : \mathbb{R}^2 \to \mathbb{P}^2$ la carta coordenada canónica como función, es decir

$$\varphi_3 (x, y) = [x : y : 1] \, .$$

Análogamente, sean $\varphi_2, \varphi_1 : \mathbb{R}^2 \to \mathbb{P}^2$ las cartas coordenadas de la segunda y primera variable respectivamente. ¿Cúal es la expresión analítica de $\varphi_2^{-1} \circ \varphi_3$? ¿En dónde está definida? A estas funciones se les llama *cambio de coordenadas*, pues en un atlas indican cómo interpretar puntos en una carta (mapa) dentro de otra.

EJERCICIO 6.33 En la carta afín canónica localiza los puntos $[2 : 0 : 1]$, $[3 : 0 : 1]$, $[0 : 2 : 1]$, $[1 : 0 : 2]$ y $[1 : 1 : -1]$. (Recuerda la sección de coordenadas baricéntricas).

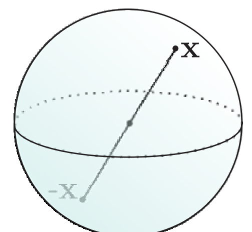
EJERCICIO 6.34 En la carta coordenada canónica, ¿cuál es la recta al infinito de la carta afín canónica?

6.5.3 La esfera módulo antípodas

Otra manera de "ver" al plano proyectivo es usando a la esfera unitaria $\mathbb{S}^2$. Puesto que cada recta por el origen en $\mathbb{R}^3$ intersecta a la esfera en dos puntos antípodas, a saber,

$$[\mathbf{x}] \cap \mathbb{S}^2 = \left\{ |\mathbf{x}|^{-1} \mathbf{x}, - |\mathbf{x}|^{-1} \mathbf{x} \right\},$$

entonces la función



$$\begin{array}{rcl} \mathbb{S}^2 & \to & \mathbb{P}^2 \\ \mathbf{x} & \mapsto & [\mathbf{x}] \end{array}$$

es sobre y 2 a 1. Es decir, la pre-imagen de cada punto consiste precisamente de dos puntos; y en nuestro caso esos dos puntos son antípodas (pues $[\mathbf{x}] = [-\mathbf{x}]$). Dicho de otra manera, la esfera unitaria $\mathbb{S}^2$ contiene representantes de todos los puntos de $\mathbb{P}^2$; más aún, contiene exactamente dos representantes de cada uno y éstos dos son antípodas. Si consideramos "ser antípodas" como una relación de equivalencia en $\mathbb{S}^2$, entonces $\mathbb{P}^2$ es el conjunto de clases de equivalencia. Esto se escribe como

$$\mathbb{P}^2 = \mathbb{S}^2 / \{\mathbf{x} \backsim -\mathbf{x}\}$$

y se lee "$\mathbb{P}^2$ es igual a $\mathbb{S}^2$ *módulo* antípodas". Más aún, como las clases de equivalencia consisten de parejas de puntos muy alejados (de hecho tan alejados como es posible), entonces en porciones pequeñas de la esfera (que no contengan una pareja de antípodas) la correspondencia es 1 a 1. Se dice entonces que la esfera es un *cubrimiento doble* del plano proyectivo.

Si con las cartas afines hicimos el símil de una fotografía fiel pero local (de un pedazo) de $\mathbb{P}^2$, ahora tendríamos que decir que este modelo es una foto global, pero sobreexpuesta (aunque con control: cada punto aparece dos veces en la foto y justo en extremos opuestos de ella).

Éste será el modelo adecuado para hablar de cuestiones métricas en el plano proyectivo. Observemos que las rectas de $\mathbb{P}^2$, al corresponder a planos por el origen de $\mathbb{R}^3$, se "levantan" a círculos máximos en $\mathbb{S}^2$. Al recorrer un círculo máximo en $\mathbb{S}^2$ para regresar al punto de partida, se recorre "abajo", en $\mathbb{P}^2$, una recta pero se le dan dos vueltas (al llegar al punto antípoda en $\mathbb{S}^2$ ya llegamos al punto de partida en $\mathbb{P}^2$ y arrancamos una nueva vuelta). Cada círculo máximo se mapea 2 a 1 a su recta correspondiente en el proyectivo. Tenemos entonces que las maneras más económicas (o rápidas) de viajar en $\mathbb{S}^2$, que son los círculos máximos, corresponden a las rectas de $\mathbb{P}^2$. Lo cual indica que debe haber nociones de distancia y ángulo en el plano proyectivo si los copiamos de su cubierta doble. Esto se hará con detalle más adelante; por lo pronto queremos dejar la idea de que hay un proyectivo rígido, con nociones métricas y que localmente parece un cascarón esférico.

6.5.4 Modelo del disco

Del modelo de la esfera podemos obtener uno nuevo como sigue. Consideremos el hemisferio superior de la esfera

$$\mathbb{S}^2_+ := \left\{(x, y, z) \in \mathbb{S}^2 \mid z \geq 0\right\}.$$

De nuevo, $\mathbb{S}^2_+$ tiene representantes de todos los puntos de $\mathbb{P}^2$, *i.e.*, la función $\mathbf{x} \mapsto [\mathbf{x}]$ restringida a $\mathbb{S}^2_+$ es sobre. Pero ahora es 2 a 1 solamente en la *frontera*, es decir, en el ecuador (cuando $z = 0$), que corresponde a la recta al infinito de nuestra carta coordenada canónica $z = 1$. Pues cada punto que no está en la frontera de $\mathbb{S}^2_+$ tiene su tercera coordenada positiva, $z > 0$, y entonces su antípoda la tiene negativa, y por tanto ya no está en $\mathbb{S}^2_+$. Si proyectamos ortogonalmente al hemisferio superior sobre el plano $z = 0$, obtenemos una biyección con el disco unitario en $\mathbb{R}^2$,

$$\mathbb{D}^2 := \left\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\right\},$$

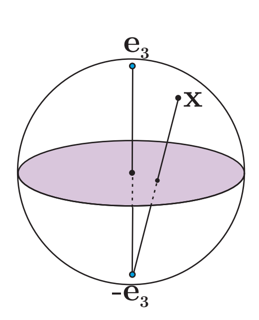
cuya frontera es $\mathbb{S}^1$. En topología, se usa al símbolo de "parcial", ∂, para denotar la frontera, y se lee "frontera", así que podemos escribir lo anterior:

$$\partial\,\mathbb{D}^2 = \mathbb{S}^1.$$

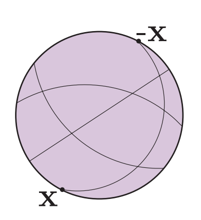
Y podemos concluir que el plano proyectivo se obtiene del disco $\mathbb{D}^2$ identificando cada punto de la frontera con su antípoda, que se escribe

$$\mathbb{P}^2 = \mathbb{D}^2 / \left\{\mathbf{x} \sim -\mathbf{x} \mid \mathbf{x} \in \mathbb{S}^1\right\}.$$

A éste lo llamaremos el *modelo ortogonal del disco*. Tiene la desventaja de que las rectas proyectivas, excepto la del infinito y las que pasan por $[\mathbf{e}_3]$, se proyectan en elipses, y que en la recta al infinito distorsiona mucho pues ahí la proyección ortogonal colapsa la dirección vertical. Esto se corrige cambiando de proyección.



La *proyección estereográfica* de la esfera consiste en proyectar desde el polo sur, $-\mathbf{e}_3$, al plano $z = 0$; es decir, a cada punto $\mathbf{x} \in \mathbb{S}^2$ lo mandamos a la intersección de su recta con $-\mathbf{e}_3$ y el plano $z = 0$. Es claro que al hemisferio superior lo manda en el disco unitario (y al inferior a su exterior, pero esto no nos interesa). Resulta que la proyección estereográfica (de $\mathbb{S}^2 - \{-\mathbf{e}_3\}$ en $\mathbb{R}^2$) preserva ángulos y manda a las líneas esféricas en círculos que intersectan a $\mathbb{S}^1$ en puntos antípodas. No demostramos lo anterior pues formalmente nunca lo necesitaremos, pero es un buen ejercicio de geometría en el espacio; aunque son las dos propiedades que hacen preferible a la proyección estereográfica sobre la ortogonal.



El *modelo del disco* (ya sin apellido, o *modelo estereográfico del disco*, si queremos ser más explícitos) se obtiene del hemisferio superior $\mathbb{S}^2_+$ proyectando estereográficamente sobre $\mathbb{D}^2$, y entonces $\mathbb{P}^2$ es este $\mathbb{D}^2$ identificando puntos antípodas en la frontera ($\{\mathbf{x} \sim -\mathbf{x} \mid \mathbf{x} \in \mathbb{S}^1\}$). Se tiene además que las rectas proyectivas son segmentos de círculo que intersectan a $\mathbb{S}^1$ en puntos antípodas (incluyendo los diámetros como círculos con centro en el infinito). Tiene la agradable particularidad de que los ángulos sí se pueden medir en él, aunque las distancias no.

EJERCICIO 6.35 En un pedazo de tela redondo ($\sim \mathbb{D}^2$), un mantelito digamos, zurce la orilla ($\partial \mathbb{D}^2$) de tal manera que cada punto quede pegado a su antípoda. ¿Se puede concluir este trabajo?

EJERCICIO 6.36 Observa que en el modelo del disco sólo se usa el plano y subconjuntos de él. Describe el análogo del modelo del disco para $\mathbb{P}^3$.

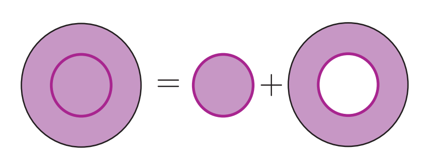
EJERCICIO 6.37 Demuestra que la proyección estereográfica $f : \mathbb{S}^2 - \{-\mathbf{e}_3\} \to \mathbb{R}^2$ tiene la fórmula analítica

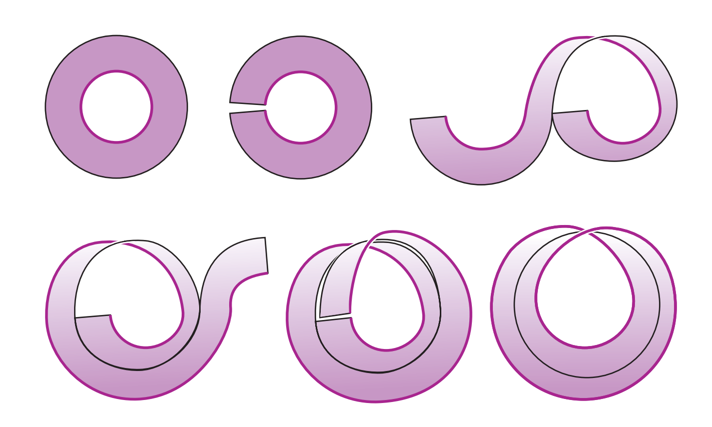
$$f(x, y, z) = \frac{1}{z+1}(x, y).$$

EJERCICIO 6.38 Demuestra que bajo la proyección estereográfica la línea esférica dada por la ecuación $ax + by + z = 0$ se mapea en el círculo con centro en (a, b) y radio $\sqrt{a^2 + b^2 + 1}$. ¿Qué líneas esféricas no caen en el enunciado anterior y a dónde se mapean?

6.5.5 La banda de Moebius más un disco

El ejercicio de zurcir el mantelito tiene respuesta negativa; no hay manera de ver al plano proyectivo como subconjunto de $\mathbb{R}^3$. Pero sí podemos verlo en dos pedazos. Si del disco $\mathbb{D}^2$ cortamos un disco interior, el de radio $1/2$ digamos, y llamémoslo $\mathbb{D}^2_{1/2}$, obtenemos un anillo y un disco (de los cuales hay que recordar que —y cómo— sus fronteras se identifican). En el anillo, además tenemos que identificar la frontera exterior antipodálmente para obtener el plano proyectivo, y esto se puede lograr haciendo un nuevo corte que nos da una banda; luego, dándole media vuelta a una mitad (estirándola), al mismo tiempo que vamos haciendo un giro (en $\mathbb{R}^3$) que pase lo de adentro hacia afuera para identificar el primer par de antípodas. Y de ahí en adelante, seguir identificando es como cerrar un *zipper*: la banda tiene que dar otro giro de π para regresar justo a pegarse con el extremo que se había quedado fijo y untado al plano. Obtenemos una banda de Moebius M. La frontera de la banda de Moebius es un círculo (se escribe, topológicamente, $\partial M = \mathbb{S}^1$) que debe identificarse con la frontera de $\mathbb{D}^2_{1/2}$ que también es un círculo. Podríamos resumir con una fórmula topológica del estilo

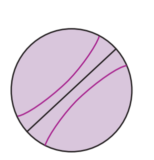


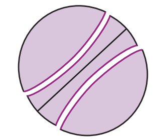


$$“\ \mathbb{P}^2 = \mathbb{D}^2 \cup_\partial M\ ”$$

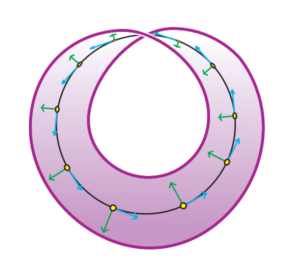
pues el disco $\mathbb{D}^2_{1/2}$ es “igual” al canónico $\mathbb{D}^2$ y tanto él como la banda de Moebius M tienen la misma frontera, $\mathbb{S}^1$, que se identifica; es una unión donde se “comparte” la frontera. El entrecomillado es porque no nos toca dar definiciones más precisas. Varias observaciones son pertinentes.

Primero, obsérvese que el círculo que originalmente era la frontera exterior del anillo viene de una línea proyectiva, la del infinito; al hacer la identificación antipodal quedó justo como una línea (punto por punto, sin redundancias, pues los puntos que debían identificarse lo hicieron), y es el corazón, alma o **ánima** de la banda de Moebius. Puesto que todas las líneas son iguales, debemos tener entonces que alrededor de cualquier línea proyectiva, su “vecindad regular”, que se obtiene alejándose un poquito en las dos direcciones normales a ella, también hay una banda de Moebius. Si hacemos esto con otra línea, un diámetro digamos en el modelo del disco, se obtiene el mismo resultado; es decir, córtese una vecindad regular de un diámetro, quedan tres pedazos, en ellos se pueden hacer las identificaciones de la frontera y nos queda, como era de esperarse, una banda de Moebius (con el diámetro como ánima) y un disco cuyas fronteras deben identificarse.





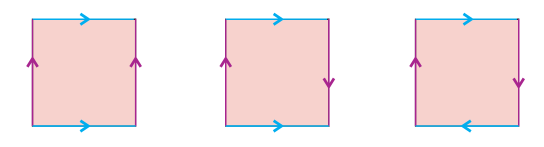
Segundo, recuérdese que en la sección última (y extra) del Capítulo 1 (si se pasó por alto, hojéese) ya habíamos obtenido una banda de Moebius abierta como el espacio de rectas en $\mathbb{R}^2$. Esa banda de Moebius abierta es justo el plano proyectivo quitándole un punto, pues las rectas en $\mathbb{R}^2$ son las proyectivas menos la del infinito, y las rectas proyectivas corresponden, por dualidad, a los puntos de $\mathbb{P}^2$; esa banda de Moebius se obtiene haciendo chiquito al disco complementario.



Y tercero: el plano proyectivo no es orientable. Si un punto viaja por el ánima de una banda de Moebius junto con una base de vectorsitos, el primero apuntando en la dirección del movimiento y el segundo perpendicular a la izquierda (lo que hemos convenido como orientación positiva), al dar una vuelta el primer vector regresa justo a su lugar, pero el segundo apunta en la dirección contraria. No hay manera de dar orientaciones locales de manera coherente en toda la banda de Moebius y por eso se dice que es no orientable. Pero esto también tiene sus ventajas, y podemos mover a la banda de Moebius dentro de sí misma —empujándola junto con el punto, trasladando en la recta que es su ánima— y al dar una vuelta completa, el efecto local es como si reflejáramos en la línea, la frontera sólo ha dado media vuelta dentro de sí misma. Esto se verá con más cuidado al estudiar las transformaciones del plano proyectivo.

EJERCICIO 6.39 [“Toro”]

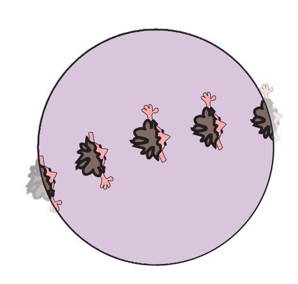
“Toro” Se pueden identificar los lados opuestos de un cuadrado de tres maneras esencialmente distintas como en la figura. La primera da la superficie (la frontera) de una dona, que en matemáticas se conoce como *toro*. La segunda da la *botella de Klein,* ¿qué da la tercera? ¿Cuáles se pueden ver en $\mathbb{R}^3$? ¿Cuáles son orientables?



6.5.6 Modelo Nintendo

Muchos juegos de computadora o Nintendo en los que el objeto que controla el jugador, una navecita digamos, se mueve libremente en dos dimensiones, tienen la particularidad de que cuando la nave llega a un extremo de la pantalla aparece justo por el lado contrario. En estos juegos, el “espacio” (virtual por necesidad) donde vive la nave es un Toro (véase el ejercicio “Toro”). Podemos entonces diseñar un juego equivalente donde el espacio sea el plano proyectivo, tomando como “cancha” un disco y haciendo que cada vez que la navecita choque con la frontera aparezca continuamente por la parte opuesta.

La navecita tiene la particularidad de ser simétrica con respecto a un eje, así que la cambiaremos por un ser bidimensional que no sea simétrico; que tenga una mano pero la otra no, un “manquito”. Si sale del centro del disco, cruza la frontera (él ni cuenta se da de que lo está haciendo) y regresa al centro, ahora está manco de la mano contraria. No se puede decir, por tanto, si es manco de la mano derecha o de la mano izquierda. Esto sucede porque el plano proyectivo no es orientable; localmente sí se puede hablar de derecha e izquierda, pero globalmente pierden su sentido. Al dar una vuelta “esencial” por su universo, el manquito ve todo como en espejo; pero no es tan grave, una vuelta más y todo regresa a la normalidad.



EJERCICIO 6.40 ¿Es orientable $\mathbb{P}^3$?

EJERCICIO 6.41 Supón que nuestro universo es no orientable. ¿Cuál sería el efecto sobre ti de un viaje intergaláctico que regrese a la Tierra pero con la orientación opuesta? ¿Sabrías leer?

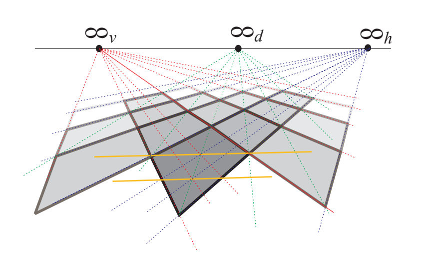
6.6 Transformaciones proyectivas

Hemos insistido en que un enfoque moderno en geometría debe incluir por fuerza un estudio de las transformaciones relevantes. Nos toca entonces estudiar las transformaciones del plano proyectivo en sí mismo que preservan la estructura que hemos definido, sus subconjuntos especiales llamados rectas.

6.6.1 Con regla

Recordemos la sección donde se construyó el método del pintor para que, con el simple uso de una regla (y lápiz y goma, por supuesto), se pudiera dibujar un piso cuadriculado en perspectiva. Ahora podemos decir mucho más cosas sobre ese método.

Si pensamos en el piso como $\mathbb{R}^2$ junto con una línea al infinito, es decir, como $\mathbb{P}^2$, y al lienzo también como $\mathbb{P}^2$ (o una porción de él), entonces estabamos rastreando, o construyendo, una función de $\mathbb{P}^2$ en sí mismo que manda rectas en rectas a partir de la imagen del cuadrado unitario o "loseta básica". La construcción se hace totalmente general cuando sabemos que cada par de rectas se intersectan, pues de las dos (imagenes de las) horizontales definimos su intersección, ∞_h; de las dos verticales el correspondiente ∞_v; a partir de estos dos puntos (que son diferentes pues las intersecciones de unas y otras son los vértices de la loseta) encontramos la imagen de la línea al infinito (que en aquel entonces llamamos horizonte), etc. Solamente se están usando los lemas de incidencia y que la función manda rectas en rectas.

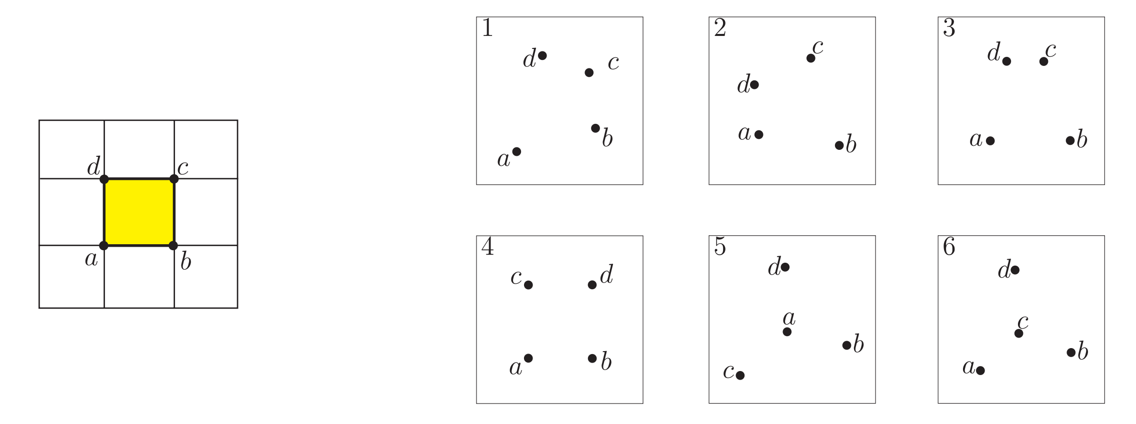


Inclusive se puede definir a dónde se mandan cuadrículas más finas al considerar la otra diagonal de la loseta, y tomar la intersección de ambas diagonales que debe ser la imagen del centro de la loseta original; a partir de este punto trazamos nuevas horizontales y verticales y el mismo proceso nos definirá a dónde van las losetitas. Para efectos de un pintor, se puede entonces aproximar (usando cuadrículas más y más finas) tanto como él quiera a una perspectiva perfecta de lo que esté dibujado en el piso. Y con variaciones o aplicaciones del mismo método a otros planos (fachadas, etc.), bastará para buenas reproducciones del mundo tridimensional sobre un lienzo. Pero para nosotros no basta.

Da la impresión de que este método tiene como corolario que para cualquier cuadrángulo (la loseta originalmente pintada y bastante arbitraria por cierto), habrá una transformación de $\mathbb{P}^2$ en $\mathbb{P}^2$ que manda rectas en rectas, y al cuadrado unitario en el dado. Este teorema es cierto, pero en su demostración por métodos sintéticos se involucran resultados fuertes, entre otros los teoremas de Desargues y de Papus (que tienen que ver con el problema de colinearidad que se planteó entonces), y también surge un problema profundo de completez. Pues para nada es claro que a cualquier punto lo podamos definir con una simple regla y un compás; tendríamos que invocar continuidad, que nos rebasa.

Pero las cosas no están tan mal como las estamos pintando. Hemos construido un poderoso método analítico con el que resolveremos las interrogantes planteadas.

EJERCICIO 6.42 Dadas las imágenes de los cuatro vértices de una loseta como en las figuras siguientes, completa a dónde tienen que ir las ocho losetas circundantes bajo una transformación proyectiva (que preserve líneas).



EJERCICIO 6.43 Añade a las figuras anteriores la horizontal y la vertical que parten a la loseta en cuatro losetitas iguales.

6.6.2 Analíticas

Una transformación de $\mathbb{P}^2$ en $\mathbb{P}^2$, pensando a sus puntos como rectas por el origen en $\mathbb{R}^3$, nos dice cómo mandar a estas rectas en ellas mismas, define una correspondencia entre ellas; si además le pedimos que mande rectas proyectivas en rectas proyectivas, la correspondencia entre rectas deberá mandar planos por el origen en planos por el origen. Y de este tipo de transformaciones en $\mathbb{R}^3$ ya hemos hablado. Tenemos las transformaciones lineales de $\mathbb{R}^3$ que cumplen justo estas propiedades, mandan al origen en sí mismo, a las rectas por él en rectas por él y a los planos por el origen en planos por el origen. ¡Ni mandadas a hacer!

Sea A una matriz invertible de 3×3, que define una transformación lineal de $\mathbb{R}^3$ en $\mathbb{R}^3$ dada por la regla de correspondencia $\mathbf{x} \mapsto A\mathbf{x}$ para todo $\mathbf{x} \in \mathbb{R}^3$. Podemos entonces definir su *transformación proyectiva asociada* como

$$\begin{aligned} f_A &: \mathbb{P}^2 \to \mathbb{P}^2 \\ f_A\left([\mathbf{x}]\right) &= [A\mathbf{x}] \,. \end{aligned}$$

Para ver que está bien definida, hay que ver que la imagen está en donde debe y que no depende del representante. Puesto que $[\mathbf{x}] \in \mathbb{P}^2$ implica que $\mathbf{x} \neq \mathbf{0}$, y esto a su vez que $A\mathbf{x} \neq \mathbf{0}$ pues escogimos a A invertible, entonces $[A\mathbf{x}] \in \mathbb{P}^2$. Además, dada $t \neq 0$, se tiene que

$$f_A\left([t\,\mathbf{x}]\right) = [A(t\,\mathbf{x})] = [t\,(A\mathbf{x})] = [A\mathbf{x}] = f_A\left([\mathbf{x}]\right).$$

Por tanto f_A está bien definida y manda rectas en rectas, pues éstas corresponden a los planos por el origen en $\mathbb{R}^3$ y A manda planos por el origen en planos por el origen.

Es facil ver que estas transformaciones forman un grupo pues las matrices invertibles (o las transformaciones lineales) lo hacen (veáse, y hágase el Ejercicio 6.44). Definimos entonces el *grupo de transformaciones proyectivas de* $\mathbb{P}^2$, o de *proyectividades*, como

$$\mathbf{Pr}(2) := \left\{ f_A : \mathbb{P}^2 \to \mathbb{P}^2 \mid A \in \mathbf{GL}(3) \right\}.$$

Por la definición, parecería que $\mathbf{Pr}(2)$ está en correspondencia con $\mathbf{GL}(3)$, pero no es así. Muchas matrices definen la misma función. Por ejemplo, si tomamos $t \neq 0$ entonces $t\,A \in \mathbf{GL}(3)$ y se tiene que

$$[(t\,A)\,\mathbf{x}] = [t\,(A\mathbf{x})] = [A\mathbf{x}]$$

para todo $[\mathbf{x}] \in \mathbb{P}^2$. Así que $f_A = f_{t\,A}$, y se valdría de nuevo hablar de $f_{[A]} := f_A$ como nos sucedió con las rectas. Lo que sí hay es una función de $\mathbf{GL}(3)$ sobre $\mathbf{Pr}(2)$, que por preservar el producto se llama *homomorfismo*. Y lo que sería bonito demostrar es que dos funciones proyectivas son iguales sólo si sus matrices son *proyectivamente equivalentes*, es decir de la forma A y $t\,A$ para $t \neq 0$. Pero esto se verá en la siguiente sección.

EJERCICIO 6.44 Demuestra que $\mathbf{Pr}(2)$ es un grupo de transformaciones.

EJERCICIO 6.45 ¿Cómo definirías $\mathbf{Pr}(3)$?

6.6.3 El Teorema de 4 en 4

Decimos que un conjunto de puntos en el plano proyectivo está en *posición general* si ninguna tercia de ellos es colineal.

Teorema 6.9 *Sean* p_0, p_1, p_2, p_3 *y* q_0, q_1, q_2, q_3 *dos cuartetas de puntos en* $\mathbb{P}^2$ *en posición general. Entonces existe una única transformación proyectiva* $f \in \boldsymbol{Pr}(2)$ *que manda una en la otra, es decir, tal que* $f(p_i) = q_i$ *para* $i = 0, 1, 2, 3$.

Demostración. Como ya hemos visto en la demostración de los teoremas análogos, bastara demostrar que una cuarteta determinada puede ir a cualquier otra (ver y hacer el Ejercicio 6.46). Como cuarteta determinada, claramente nos conviene escoger a la cuarteta canónica $[\mathbf{e}_0], [\mathbf{e}_1], [\mathbf{e}_2], [\mathbf{e}_3]$ donde $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ son la base canónica de $\mathbb{R}^3$, y $\mathbf{e}_0 = (1, 1, 1) = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$.

Sea q_0, q_1, q_2, q_3 una cuarteta de puntos cualquiera en $\mathbb{P}^2$ en posición general. Tenemos entonces que $q_i = [\mathbf{u}_i]$ para algún vector $\mathbf{u}_i \in \mathbb{R}^3$. Es muy fácil lograr que la base canónica vaya a donde debe. Sea

$$A' = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$$

la matriz 3×3 que cumple $A'e_1 = u_1$, $A'e_2 = u_2$ y $A'e_3 = u_3$. Que $[u_1]$, $[u_2]$ y $[u_3]$ no sean colineales en $\mathbb{P}^2$ implica que los vectores u_1, u_2 y u_3 no son coplanares en $\mathbb{R}^3$ y, por lo tanto, que la matriz A' es invertible. Entonces su función proyectiva asociada f' está bien definida y cumple que $f'([e_1]) = [u_1]$, $f'([e_2]) = [u_2]$ y $f'([e_3]) = [u_3]$, pero su valor en $[e_0]$ no tiene por qué ser $[u_0]$; por eso hemos usado las primas, todavía no es la buena.

Observemos que aún podemos modificar a A' de tal manera que los puntos de la base canónica vayan a dónde deben, a saber, si multiplicamos una columna por una constante no cero. Por ejemplo la matriz

$$A'' = (t\,u_1, u_2, u_3)$$

cumple que $\left[A''e_1\right] = [t\,u_1] = [u_1]$ (aunque ya no sea cierto que mande al vector e_1 en el vector u_1). De tal manera que todavía tenemos tres grados de libertad para modificar a A' multiplicando sus tres columnas por constantes distintas de cero.

Como u_1, u_2 y u_3 forman una base de $\mathbb{R}^3$, existen constantes λ_1, λ_2 y λ_3 tales que

$$u_0 = \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3.$$

Sea A (ahora sí la buena) la matriz

$$A = (\lambda_1 u_1, \lambda_2 u_2, \lambda_3 u_3)\,.$$

Puesto que $[u_0]$ no es colineal con $[u_2]$ y $[u_3]$, entonces u_0 no es coplanar con u_2 y u_3, y por lo tanto $\lambda_1 \neq 0$. Análogamente, $\lambda_2 \neq 0$ y $\lambda_3 \neq 0$, y por tanto $\det(A) = \lambda_1\lambda_2\lambda_3 \det(A') \neq 0$. Lo cual nos dice que $A \in \mathbf{GL}(3)$ y tiene una transformación proyectiva asociada $f = f_A$ que cumple

$$\begin{aligned} f([e_0]) &= [Ae_0] = [\lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3] = [u_0] \\ f([e_1]) &= [Ae_1] = [\lambda_1 u_1] = [u_1] \\ f([e_2]) &= [Ae_2] = [\lambda_2 u_2] = [u_2] \\ f([e_3]) &= [Ae_3] = [\lambda_3 u_3] = [u_3] \end{aligned}$$

y por tanto demuestra la existencia.

Para demostrar la unicidad, de nuevo (por las propiedades de grupo) basta demostrar un caso particular (ver y hacer el Ejercicio 6.47). Supongamos que $f \in \mathbf{Pr}(2)$ es tal que $f([e_i]) = [e_i]$ para $i = 0, 1, 2, 3$. Sabemos que f está dada por una matriz

$$A = (v_1, v_2, v_3)\,.$$

Puesto que $[e_1] = f([e_1]) = [Ae_1] = [v_1]$, entonces existe $t_1 \neq 0$ tal que $v_1 = t_1 e_1$; y análogamente existen $t_2 \neq 0$ y $t_3 \neq 0$ tales que $v_2 = t_2 e_2$ y $v_3 = t_3 e_3$; es decir, A es una matriz diagonal. Pero se tiene además que

$$[1:1:1] = [e_0] = f([e_0]) = [Ae_0] = [t_1 : t_2 : t_3]\,,$$

lo cual implica que $t_1 = t_2 = t_3 = t$. Por tanto

$$A = \begin{pmatrix} t & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & t \end{pmatrix} = t\,I$$

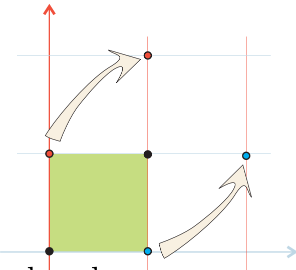
y esto, como ya hemos visto, implica que $f = f_I = \mathrm{id}_{\mathbb{P}^2}$. □

De este teorema y su demostración obtendremos, por lo pronto, dos implicaciones. La primera es que el método del pintor que hemos descrito es coherente. Pues si las cuatro esquinas de la loseta inicial están en posición general, entonces existe una transformación proyectiva, dada analíticamente, que manda a todo $\mathbb{P}^2$ en sí mismo y que manda rectas en rectas (preservando por tanto colinearidad y concurrencia), de tal manera que el método constructivo que describimos es obligado.

La otra implicación es concluir la descripción del grupo $\mathbf{Pr}(2)$ en términos de matrices. Si declaramos que dos matrices A y B son equivalentes cuando existe $t \neq 0$ tal que $B = t\,A$ y denotamos con $[A]$ a la clase de equivalencia de A, entonces las transformaciones proyectivas están en correspondencia natural y biyectiva con el conjunto de clases de equivalencia de matrices invertibles, es decir, con $\{[A] \mid A \in \mathbf{GL}(3)\}$; que a veces se llama $\mathbf{GL}(3)$ proyectivizado y se le denota $\mathbf{PGL}(3)$.

Ejemplo

Para fijar ideas, veamos un ejemplo concreto. Queremos expresar analíticamente la transformación proyectiva de $\mathbb{R}^2$ en $\mathbb{R}^2$ (pensados como la carta coordenada canónica de $\mathbb{P}^2$) que hace lo siguiente



$$\begin{array}{ccc} (0,0) & \mapsto & (0,0) \\ (1,0) & \mapsto & (2,1) \\ (0,1) & \mapsto & (1,2) \\ (1,1) & \mapsto & (1,1)\,. \end{array}$$

Proyectivamente, a cada punto hay que agregarle un 1 a sus coordenadas para representarlos con un vector en $\mathbb{R}^3$, pues nos dieron sólo datos finitos, y el problema consiste en encontrar una matriz A de 3×3 que mande a cada vector en la línea generada por el vector correspondiente. Si llamamos $\mathbf{a}_i$ $(i = 0, 1, 2, 3)$ a los correspondientes vectores de la izquierda y $\mathbf{b}_i$, respectivamente, a los que tienen que ser sus imágenes, necesitamos encontrar A tal que $A\mathbf{a}_i = \lambda_i \mathbf{b}_i$ para algunas $\lambda_i \neq 0$. Puesto que A tiene 9 entradas incógnitas y hemos agregado las λ_i, tenemos 13 incógnitas en total; cada ecuación $A\mathbf{a}_i = \lambda_i \mathbf{b}_i$ representa tres ecuaciones lineales donde los datos dados proveen las constantes. Así que tenemos 12 ecuaciones lineales con 13 incógnitas que, en principio, debían tener soluciones a un parámetro, que corresponde al hecho de que si A es solución tambien lo es tA para $t \neq 0$. Aunque no es tan difícil como parece, pues se pueden reagrupar las ecuaciones para hacerse similares por grupos, mejor intentemos otro método: siguiendo la demostración del teorema.

Sea $\mathfrak{E}$ la cuarteta canónica $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ que usamos en el teorema ($\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ es la base canónica de $\mathbb{R}^3$ y $\mathbf{e}_0 = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$). Nos conviene reenumerar nuestros datos del dominio para que se parezcan más, de entrada, a $\mathfrak{E}$; sean entonces $\mathbf{a}_3 = (0, 0, 1)$, $\mathbf{a}_1 = (1, 0, 1)$, $\mathbf{a}_2 = (0, 1, 1)$ y $\mathbf{a}_0 = (1, 1, 1)$ y denotemos con $\mathfrak{A}$ a la cuarteta ordenada $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$. Respectivamente, sea $\mathfrak{B}$ la cuarteta $\mathbf{b}_0 = (1, 1, 1)$, $\mathbf{b}_1 = (2, 1, 1)$, $\mathbf{b}_2 = (1, 2, 1)$ y $\mathbf{b}_3 = (0, 0, 1)$. Nuestro problema se parte en dos: encontrar una matriz A que mande a la cuarteta $\mathfrak{E}$ en la $\mathfrak{A}$ (ojo, esta matriz no es la del parrafo anterior) y una matriz B que mande a $\mathfrak{E}$ en $\mathfrak{B}$. Para, finalmente, obtener la solución BA^{-1} que manda a las cuartetas $\mathfrak{A}$ en $\mathfrak{B}$:

$$\mathfrak{A} \xleftarrow{A} \mathfrak{E} \xrightarrow{B} \mathfrak{B}.$$

La matriz $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ manda a $\mathbf{e}_i$ en $\mathbf{a}_i$ para $i = 1, 2, 3$; pero sólo nos interesa que mande a $[\mathbf{e}_i]$ en $[\mathbf{a}_i]$ (los puntos proyectivos o las líneas), y esto lo logran todas las matrices $(\lambda_1 \mathbf{a}_1, \lambda_2 \mathbf{a}_2, \lambda_3 \mathbf{a}_3)$ con $\lambda_i \neq 0$. Así que nos interesa una solución de la ecuación

$$(\lambda_1 \mathbf{a}_1, \lambda_2 \mathbf{a}_2, \lambda_3 \mathbf{a}_3)\ \mathbf{e}_0 = \lambda_0 \mathbf{a}_0$$

para atinarle al cuarto punto. Al sustituir nuestros datos tenemos el sistema

$$\begin{array}{ccccccc} \lambda_1 & & & & & = & \lambda_0 \\ & & \lambda_2 & & & = & \lambda_0 \\ \lambda_1 & + & \lambda_2 & + & \lambda_3 & = & \lambda_0. \end{array}$$

cuya solución más bonita es $\lambda_0 = \lambda_1 = \lambda_2 = 1$ y $\lambda_3 = -1$; que nos da como matriz que manda a $\mathfrak{E}$ en $\mathfrak{A}$:

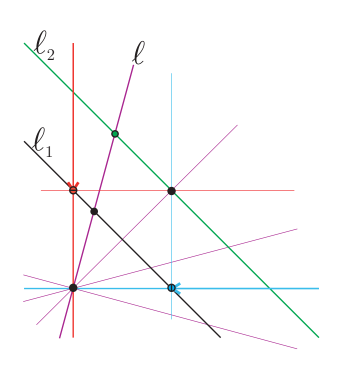
$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix}$$

Ya que necesitaremos sacar su inversa, de una vez: resulta que es A, es decir, $A^2 = I$ (hágase mentalmente el cálculo). Es, en cierta manera, una reflexión y vale la pena analizar su comportamiento geométrico antes de proseguir con el ejemplo.

La matriz A manda a la cuarteta canónica $\mathfrak{E}$ en la "loseta canónica" $\mathfrak{A}$ y, puesto que $A^2 = I$, A manda entonces a la loseta canónica en la cuarteta $\mathfrak{E}$; es decir, las intercambia. Su expresión analítica como función de $\mathbb{R}^2$ en $\mathbb{R}^2$ es

$$f(x, y) = \left(\frac{x}{x + y - 1}, \frac{y}{x + y - 1} \right)$$

(pues se añade un 1, se multiplica por A y se divide entre la tercera coordenada), de donde podemos sacar más información o corroborarla. Notemos primero que sus



puntos fijos son el punto $(0,0) = \mathbf{a}_3$ y la recta $\ell_2 : x + y = 2$ (que pasa por $\mathbf{a}_0$), pues son las únicas soluciones de las ecuaciones $f(x,y) = (x,y)$. La recta $\ell_1 : x + y = 1$ va a la recta al infinito ℓ_∞ (que ya sabíamos pues $\mathbf{a}_1$ y $\mathbf{a}_2$ que la generan lo hacen), pero más aún, se intercambia con ella.

Entonces, si ℓ es cualquier recta que pasa por el origen tenemos que f la manda en sí misma por medio de la reflexión proyectiva en los puntos $\ell \cap \ell_2$ y $\mathbf{0}$; pues estos se quedan fijos y el punto medio de su segmento, $\ell \cap \ell_1$, se intercambia con su punto al infinito $\ell \cap \ell_\infty$. Podríamos decir que f, o equivalentemente A, es *la reflexión en la recta* ℓ_2 *respecto al punto* $\mathbf{a}_3$ $(= \mathbf{0})$; y esto tiene sentido para cualquier recta y un punto fuera de ella. Por ejemplo, si tomamos cualquier recta y el punto al infinito en su dirección ortogonal, esta reflexión es justo la euclidiana (isométrica), y para cualquier otro punto al infinito es una reflexión afín que para reflejar usa una dirección distinta de la ortogonal.

Pero si vemos a f con lupa, muy cerca de $\mathbf{0}$, ahí parece una rotación de ángulo π, pues todas las rectas que pasan por él únicamente cambian de dirección. Y esto también es cierto; si tomamos la reflexión en la recta al infinito repecto a cualquier punto $\mathbf{p}$, cada recta que pasa por $\mathbf{p}$ se refleja en $\mathbf{p}$, simplemente cambia rígidamente de dirección, y entonces es la rotación isométrica por un ángulo π alrededor de $\mathbf{p}$.

Después de esta digresión respecto a la matriz A, regresemos a nuestro problema. Nos faltaba encontrar la matriz B que mande a la cuarteta canónica en la $\mathfrak{B}$. Pero ahora al primer tiro cae el gol: la matriz

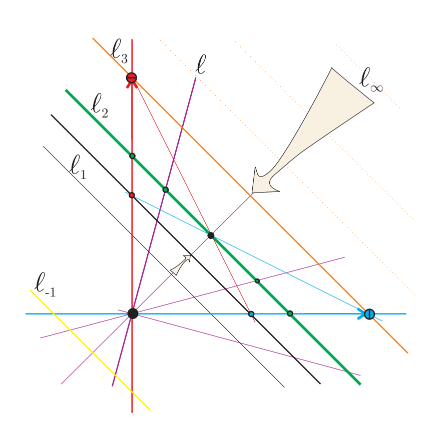
$$B = (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

cumple que $B\mathbf{b}_0 = 3\mathbf{b}_0$.

Para entender mejor el efecto geométrico de lo que hace B nos conviene partirla en dos movimientos. El primero, que fijando al origen y todas sus líneas (como líneas, no punto a punto), trae, jala, a la línea al infinito hasta la línea $\ell_3 : x+y = 3$, fijando punto a punto a la línea ℓ_2. Entonces en cada línea ℓ que pasa por el origen tiene que ser una expansión de $\mathbf{0}$ a $\ell \cap \ell_2$ y consecuentemente empuja a la línea ℓ_1 hacia ℓ_2 (y expande a cualquier segmento de $\mathbf{0}$ a ℓ_2 hacia ℓ_2). Esta transformación, donde $\mathbf{0}$ sigue "viendo" exactamente lo mismo, es una suerte de homotecia desde el origen con horizonte en ℓ_2, y tiene la matriz

$$B_1 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Y la línea que va a cubrir la línea al infinito es $\ell_{-1} : x+y = -1$.



Y después hay que componerla con una simple transformación líneal B_2 que acerca a los dos ejes para llevarlos a las líneas por $(2,1)$ y $(1,2)$. (esta transformación ya deja al infinito en su lugar):

$$B = B_2 B_1 = \frac{1}{3}\begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

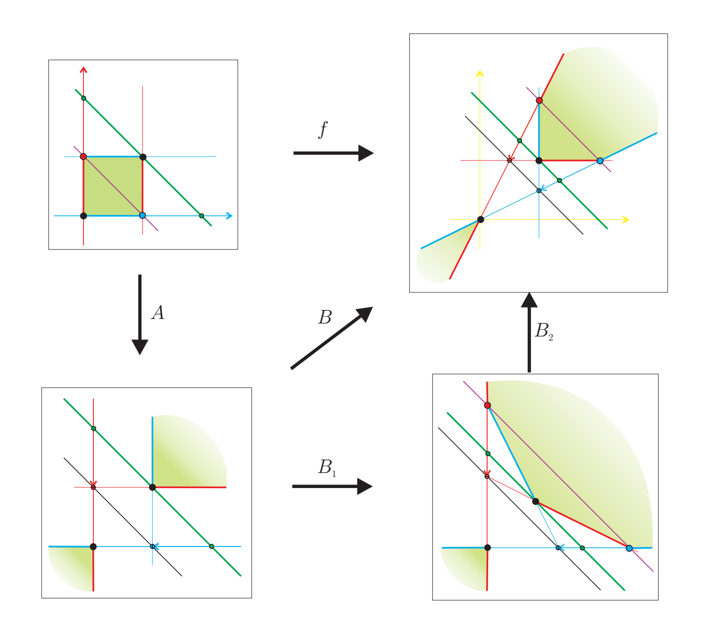
La matriz que buscabamos es entonces

$$BA = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 2 & 2 & -1 \end{pmatrix}.$$

Lo cual en coordenadas de la carta canónica, nos da la función

$$f(x,y) = \frac{1}{2x+2y-1}(2x+y, x+2y).$$

En la siguiente figura se persigue a la "loseta" canónica bajo las transformaciones proyectivas que acabamos de usar en el ejemplo.



EJERCICIO 6.46 Demuestra que si existe $f \in \mathbf{Pr}(2)$ tal que $f(a_i) = q_i$ para $i = 0,1,2,3$ para cualquier cuarteta q_0, q_1, q_2, q_3 de puntos en $\mathbb{P}^2$ en posición general, donde a_0, a_1, a_2, a_3 es una cuarteta fija de puntos en $\mathbb{P}^2$ en posición general, entonces se sigue la parte de existencia del Teorema 6.9

EJERCICIO 6.47 Demuestra que si es cierto que dada $f \in \mathbf{Pr}(2)$ tal que $f([\boldsymbol{e}_i]) = [\boldsymbol{e}_i]$ para $i = 0, 1, 2, 3$ implica que $f = \mathrm{id}_{\mathbb{P}^2}$, entonces se sigue la parte de unicidad del Teorema 6.9.

EJERCICIO 6.48 Sean $\boldsymbol{e}_1 = [1 : 0 : 0]$, $\boldsymbol{e}_2 = [0 : 1 : 0]$, $\boldsymbol{e}_3 = [0 : 0 : 1]$ y $\boldsymbol{e}_0 = [1 : 1 : 1]$. Encuentra explícitamente las transformaciones proyectivas que mandan a estos cuatro puntos respectivamente en:

a) $[1 : 0 : 0]$, $[0 : 1 : 0]$, $[0 : 0 : 1]$, $[1 : 1 : -1]$.
b) $[0 : 1 : 1]$, $[1 : 0 : 1]$, $[1 : 1 : 0]$, $[1 : 1 : 1]$.
c) $[1 : 2 : 2]$, $[-1 : 1 : 2]$, $[0 : 2 : 1]$, $[-2 : -1 : 3]$.

EJERCICIO 6.49 En la carta afín canónica $(x + y + z = 1)$ dibuja la imagen de la línea al infinito bajo las transformaciónes proyectivas del ejercicio anterior.

EJERCICIO 6.50 En la carta coordenada canónica $(z = 1)$ dibuja la imagen de la línea al infinito bajo las transformaciónes proyectivas del Ejercicio 6.48.

6.6.4 Transformaciones afines

Puesto que $\mathbb{R}^2$ se puede ver de manera canónica dentro de $\mathbb{P}^2$ como la carta euclidiana distinguida ($(x, y) \mapsto [x : y : 1]$), es natural preguntarse si los grupos de transformaciones de $\mathbb{R}^2$ que ya conocemos se pueden ver dentro del grupo de transformaciones proyectivas. La respuesta es muy bonita. Sea $\ell_\infty = \ell_{[\boldsymbol{e}_3]}$ la recta al infinito dada por la ecuación $z = 0$ en $\mathbb{P}^2$, y sea G el conjunto de transformaciones proyectivas que dejan a la recta al infinito en su lugar (la pueden mover punto a punto pero no como conjunto). Esto es,

$$G = \{f \in \mathbf{Pr}(2) \mid f(\ell_\infty) = \ell_\infty\}.$$

Entonces G es un subgrupo que se identifica naturalmente con el grupo de transformaciones afines $\mathbf{Af}(2)$. Es decir, las transformaciones proyectivas que dejan al infinito en el infinito son precisamente las afines (como en el caso de dimensión 1). Demostrémoslo.

Sea $A \in \mathbf{GL}(3)$ una matriz que representa a $f \in G$ (A está bien definida salvo constantes no cero). Puesto que $[\boldsymbol{e}_1] = [1 : 0 : 0] \in \ell_\infty$ entonces $f([\boldsymbol{e}_1]) = [A\boldsymbol{e}_1] \in \ell_\infty$; lo cual implica que la tercera coordenada de $A\boldsymbol{e}_1$, que es la primera columna de A, es cero. Análogamente, con $\boldsymbol{e}_2$, se obtiene que la segunda columna de A tiene la tercera coordenada igual a cero. Si la tercera coordenada de la tercera columna también fuera cero, A no sería invertible; lo cual implica (multiplicando por la constante adecuada) que podemos suponer que es precisamente 1. Entonces A tiene la forma

$$A = \begin{pmatrix} a & b & e \\ c & d & f \\ 0 & 0 & 1 \end{pmatrix}.$$

Esta matriz deja fijo (como conjunto) al plano canónico $z = 1$, pues

$$A \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} ax + by + e \\ cx + dy + f \\ 1 \end{pmatrix},$$

que claramente corresponde (pensando a $\mathbb{R}^2$ como el plano canónico) a la función de $\mathbb{R}^2$ en $\mathbb{R}^2$:

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}.$$

Sea A' la matriz 2×2 anterior (el bloque superior izquierdo de A), entonces $\det(A') = \det(A) \neq 0$ y por tanto esta función es una transformación afín.

Inversamente, para cualquier transformación afín de $\mathbb{R}^2$ se puede construir una mátriz de 3×3 como arriba (que resulta ser invertible) y la transformación proyectiva correspondiente dejará fija a la recta al infinito (pues manda a $[\boldsymbol{e}_1]$ y a $[\boldsymbol{e}_2]$ en puntos de esa recta). Como la recta $z = 0$ es igual a cualquier otra, hemos demostrado el siguiente teorema.

Teorema 6.10 *Dada cualquier recta ℓ de $\mathbb{P}^2$, hay un isomorfismo de grupos*

$$\boldsymbol{Af}(2) \simeq \{f \in \boldsymbol{Pr}(2) \mid f(\ell) = \ell\},$$

es decir, las transformaciones proyectivas que fijan una recta se puede identificar con las transformaciones afines de la carta afín correspondiente $\mathbb{P}^2 - \ell$.

EJERCICIO 6.51 Sea G un grupo de transformaciones de un conjunto X. Demuestra que para cualquier subconjunto $Y \subset X$, el conjunto de transformaciones que fijan a Y,

$$G_Y = \{f \in G \mid f(Y) = Y\},$$

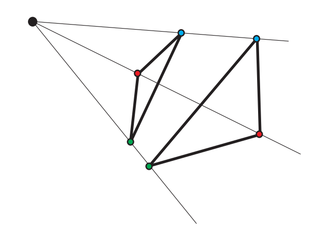
es un subgrupo de G.

EJERCICIO 6.52 En la demostración que dimos ¿a qué matrices de 3×3 corresponden las transformaciones lineales de $\mathbb{R}^2$?

EJERCICIO 6.53 Considera la carta coordenada canónica $\mathbb{A}_z$ definida por $z = 1$ con coordenadas naturales (x, y). ¿Cuáles transformaciones proyectivas del Ejercicio 6.48 se ven como transformaciones afines en esta carta?. Descríbelas geométricamente.

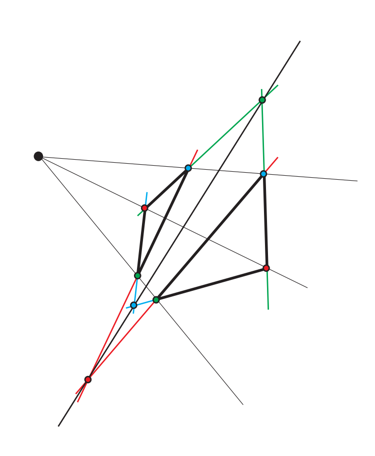
6.6.5 *Teoremas de Desargues y de Papus

En esta sección demostramos dos teoremas clásicos que precedieron, como teoría aceptada, a la geometría proyectiva y que entonces la invocan como el hábitat para su planteamiento elegante y natural.



Desargues es considerado el fundador de la geometría proyectiva como estudio sistemático de la "perspectiva". El teorema que lleva su nombre se refiere a parejas de triángulos en perspectiva. Se dice que dos triángulos *están en perspectiva* si hay una correspondencia entre sus vértices de tal manera que las tres líneas que unen vértices correspondientes son concurrentes. Al punto de concurrencia se le llama el *foco*, y desde él veríamos un solo triángulo (de ahí el nombre de "en perspectiva").

Teorema 6.11 (Desargues) *Si dos triángulos están en perspectiva, entonces los puntos de intersección de sus lados correspondientes son colineales.*



Aunque no hayamos especificado que los dos triángulos viven en el plano proyectivo, el enunciado del teorema en sí lo exige, pues habla de los puntos de intersección como un hecho. El enunciado en el caso euclidiano (recordemos que Desargues no sabía aún del plano proyectivo) debería tener varios apartados por si alguna pareja, o algunas, de lados correspondientes fueran paralelas, y también podría enunciarse para triángulos en perspectiva desde el infinito. De tal manera que, por la posibilidad euclidiana del paralelismo, en términos euclidianos clásicos este teorema se convierte en una colección de enunciados similares con pruebas similares. Desde Desargues, y él sí lo notó, fue un teorema en busca de un contexto amplio y general; lo encontró en la geometría proyectiva.

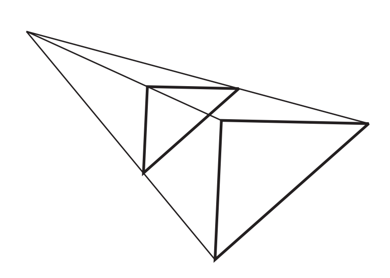
Otro aspecto interesante del Teorema de Desargues es que si se le piensa en dimensión 3 y no 2 (nótese que en el enunciado nunca especificamos la dimensión), su demostración es bastante fácil pensando, de nuevo, en términos proyectivos (sin excepciones). En el caso en que los triángulos no están en un plano, la hipótesis de que están en perspectiva es que sus lados correspondientes son segmentos en las caras de un cono triangular, y entonces las líneas correspondientes se intersectan por estar en un plano (determinado por un lado del cono). Considérense los dos planos generados por los triángulos y sea ℓ su línea de intersección: estos planos intersectan a cada lado del cono en los lados correspondientes de los triángulos y entonces ℓ los intersecta en su punto de intersección. Así que ℓ es la línea deseada. ...Dibujo

Faltaría el caso en el que todos los planos que se usaron en el argumento anterior son uno solo, y podría argumentarse por continuidad "levantando" a la tercera dimensión y luego volviendo a proyectar. Pero mejor seguiremos un camino intrínseco que ilustra el juego de la dualidad.

Recordemos el principio de dualidad que nos permite traducir cualquier enunciado o demostración cambiando recta por punto y "estar en" por "pasar por" (o bien usar únicamente el concepto incidencia). El Teorema de Desargues es justo un enunciado sobre rectas y puntos y su relación de incidencia. Más aún, el dual de su hipótesis es

precisamente su conclusión. El dual de la definición de estar en perspectiva es: se dice que dos triángulos están en *perspectiva dual* si hay una correspondencia entre sus lados de tal manera que los tres puntos de intersección de los lados correspondientes son colineales; la recta que los contiene se llama *focal.* De tal modo que lo que Desargues dice es que si dos triángulos están en perspectiva entonces están en perspectiva dual. Lo que demostraremos es su dual.

Teorema 6.12 (Desargues dual) *Si dos triángulos en* $\mathbb{P}^2$ *están en perspectiva dual entonces están en perspectiva.*



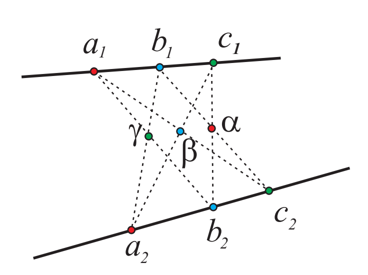
Demostración. Ambas condiciones se preservan por transformaciones proyectivas. Por tanto podemos suponer que la recta focal de la perspectiva dual, ℓ, está en el infinito del plano euclidiano. Entonces la hipótesis da dos triángulos cuyos lados son paralelos. Si son del mismo tamaño, están en perspectiva desde un punto al infinito. Si no, hay una homotecia que manda a uno en el otro y están en perspectiva desde el centro de esa homotecia. (Esto último también se puede razonar analíticamente o por semejanza de triángulos, Ejercicio 6.54). □

EJERCICIO 6.54 Sean T_1 y T_2 triángulos en $\mathbb{R}^2$ con vértices a_1, b_1, c_1 y a_2, b_2, c_2 respectivamente, y tales que $\langle a_1, b_1\rangle \parallel \langle a_2, b_2\rangle$, $\langle b_1, c_1\rangle \parallel \langle b_2, c_2\rangle$ y $\langle c_1, a_1\rangle \parallel \langle c_2, a_2\rangle$; donde $\langle p, q\rangle$ denota la recta por p y q. Demuestra, sin usar transformaciones, que entonces están en perspectiva, es decir, que las líneas $\langle a_1, a_2\rangle$, $\langle b_1, b_2\rangle$ y $\langle c_1, c_2\rangle$ son concurrentes

---

Papus

En cuanto a su historia, el Teorema de Papus es análogo al de Desargues. Aunque su enunciado es un poco más sofisticado, vivió un buen tiempo como teorema en busca de un contexto amplio; se refiere a los puntos que uno puede construir a partir de dos ternas colineales.



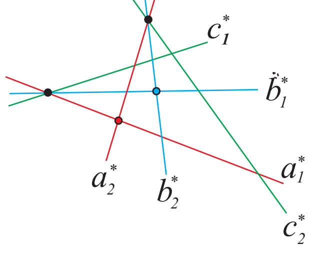
Teorema 6.13 (Papus) *Sean* a_1, b_1, c_1, a_2, b_2, c_2 *seis puntos distintos tales que* a_1, b_1, c_1 *y* a_2, b_2, c_2 *son colineales pero en rectas distintas, entonces su* terna de Papus

$$\begin{aligned}\alpha &= \langle b_1, c_2\rangle \cap \langle b_2, c_1\rangle\\ \beta &= \langle c_1, a_2\rangle \cap \langle c_2, a_1\rangle\\ \gamma &= \langle a_1, b_2\rangle \cap \langle a_2, b_1\rangle\end{aligned}$$

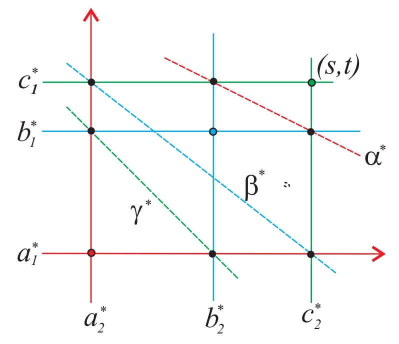
es colineal.

Antes de demostrarlo, notemos que su enunciado es netamente proyectivo. Su conclusión es sobre puntos de intersección de rectas, y si nos pusiéramos euclidianos, sería muy engorroso enunciar los distintos casos especiales de paralelismo tanto en la hipótesis como en la conclusión.

Demostración. Veamos el caso en que ninguno de los puntos es la intersección de las dos rectas. Consideremos la configuración dual; es decir, sean a_i^*, b_i^*, c_i^* las rectas polares de a_i, b_i, c_i respectivamente, para $i = 1, 2$. Se tiene entonces que cada terna de rectas es concurrente.



Mediante una transformación proyectiva, y quizá renombrando los puntos, podemos mandar a



$$\begin{array}{rcl} a_1^* \cap b_1^* & \mapsto & [1:0:0] \\ a_2^* \cap b_2^* & \mapsto & [0:1:0] \\ a_1^* \cap a_2^* & \mapsto & [0:0:1] \\ b_1^* \cap b_2^* & \mapsto & [1:1:1]\,. \end{array}$$

Para entonces trabajar en el $\mathbb{R}^2$ de la carta canónica. Como c_1^* y c_2^* no son la recta al infinito, quedan determinadas por su punto de intersección, $(s,t) = c_1^* \cap c_2^*$. Entonces las rectas polares a la terna de Papus tienen las ecuaciones

$$\begin{array}{rcc} \gamma^* & : & x + y = 1 \\ \beta^* & : & t\,x + s\,y = ts \\ \alpha^* & : & (t-1)\,x + (s-1)\,y = ts - 1, \end{array}$$

que indican claramente que son rectas concurrentes, y por tanto sus puntos polares son colineales.

El caso faltante, y posibles degeneraciones de las hipótesis, se dejan al lector. □

EJERCICIO 6.55 Demuestra el Teorema de Papus cuando uno de los puntos es la intersección de las dos rectas.

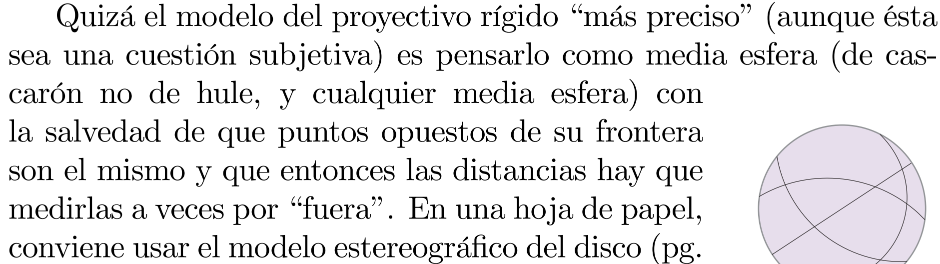
EJERCICIO 6.56 ¿Se puede relajar el teorema a que los seis puntos no sean necesariamente distintos? ¿A qué casos?

6.7 El plano proyectivo rígido

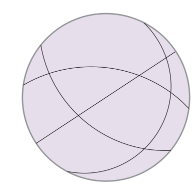
Hasta ahora sólo hemos considerado el plano proyectivo "*holgado*", con su estructura de incidencia (las rectas como subconjuntos distinguidos) y las transformaciones que la preservan (las proyectividades). Pero también tiene nociones métricas naturales (distancia y ángulos) y con ellas se convierte en el plano proyectivo "*rígido*" también

conocido como "*plano elíptico*" o podríamos decir que es el "*plano esférico*"; está más en la idea general de "geometría" que tenían los griegos y entonces sí es un contraejemplo de la necesidad del Quinto Postulado: cumple todos los axiomas de Euclides excepto el Quinto.

Como habíamos anticipado en la Sección 6.5.3, las nociones de distancia y ángulo en $\mathbb{P}^2$ se heredan naturalmente de la esfera $\mathbb{S}^2$. Pues ahí vimos que $\mathbb{S}^2$ cubre 2 a 1 a $\mathbb{P}^2$ (este último se obtiene identificando antipodas en $\mathbb{S}^2$) y localmente (es decir, en pequeños pedazos) son idénticos.



Quizá el modelo del proyectivo rígido "más preciso" (aunque ésta sea una cuestión subjetiva) es pensarlo como media esfera (de cascarón no de hule, y cualquier media esfera) con la salvedad de que puntos opuestos de su frontera son el mismo y que entonces las distancias hay que medirlas a veces por "fuera". En una hoja de papel, conviene usar el modelo estereográfico del disco (pg. 279) pues tiene la gracia de preservar ángulos (aunque no distancias); y, por supuesto, puntos antípodas en la frontera se identifican.



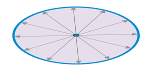
Más formalmente, sean $[\mathbf{x}]$ y $[\mathbf{y}]$ dos puntos en $\mathbb{P}^2$. Como rectas en $\mathbb{R}^3$ generan un plano (euclidiano) y en él estas dos rectas forman dos ángulos; definimos su *distancia* $\mathrm{d}([\mathbf{x}],[\mathbf{y}])$ como el mínimo de ellos. De tal manera que

$$0 \leq \mathrm{d}([\mathbf{x}],[\mathbf{y}]) \leq \pi/2\ ;$$

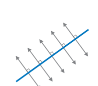
la igualdad con 0 se da sólo cuando $[\mathbf{x}] = [\mathbf{y}]$, y la cota máxima se alcanza cuando las rectas $[\mathbf{x}]$ y $[\mathbf{y}]$ son perpendiculares, es decir, cuando $[\mathbf{y}] \in \ell_{[\mathbf{x}]}$. Si se quiere una fórmula más analítica, podemos definir

$$\mathrm{d}([\mathbf{x}],[\mathbf{y}]) := \arccos\left(\left|\frac{\mathbf{x}\cdot\mathbf{y}}{|\mathbf{x}|\,|\mathbf{y}|}\right|\right). \qquad (6.15)$$

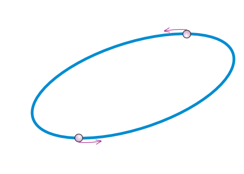
Las curvas en $\mathbb{P}^2$ cuyo recorrido logra las distancias (las de distancia mínima) o "*geodésicas*" son justo las rectas proyectivas (que corresponden a planos por 0 en $\mathbb{R}^3$, o bien a los círculos máximos en $\mathbb{S}^2$). Todas las rectas proyectivas tienen longitud π (o "circunferencia" podría decirse, pues son topológicamente círculos) y, por eso, lo más alejados que pueden estar dos puntos en una de ellas es la mitad cuando ambos segmentos de uno a otro miden lo mismo.

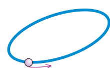


La recta polar de un punto $[\mathbf{x}]$, $\ell_{[\mathbf{x}]}$, consta entonces de todos los puntos a distancia máxima de $[\mathbf{x}]$, los más alejados de él; y para cada uno de ellos, $[\mathbf{y}] \in \ell_{[\mathbf{x}]}$ digamos, se tienen justo dos geodésicas para llegar de $[\mathbf{x}]$ a él, que juntas forman una recta. E inversamente, para cada recta ℓ hay un único punto, su polar, que está a distancia $\pi/2$ de todos sus puntos, y se llega a él saliendo *ortogonalmente* hacia cualquiera de los dos lados y desde cualquier punto de ℓ ... pero nos estamos adelantando.



La noción de ángulo en $\mathbb{P}^2$ también se copia de $\mathbb{S}^2$. Notemos que si bien las rectas de $\mathbb{P}^2$ corresponden biunívocamente a los puntos de $\mathbb{P}^2$, tenemos una correspondencia natural entre las rectas orientadas de $\mathbb{P}^2$ y los puntos de la esfera $\mathbb{S}^2$. Pues la *orientación* de una recta en $\mathbb{P}^2$ (una partícula que viaja inercialmente en ella, puede pensarse) se levanta naturalmente a una orientación de su círculo máximo correspondiente (dos partículas antípodas persiguiéndose inercial e inútilmente); y, como vimos en la Sección 5.2.1, a esta recta esférica orientada le corresponde un punto polar en $\mathbb{S}^2$. Los *ángulos* entre rectas orientadas en $\mathbb{P}^2$ se miden entonces como sus correspondientes en $\mathbb{S}^2$ (por la distancia de sus puntos polares).





Cada uno de los teoremas o resultados métricos que demostramos para la esfera (en las Secciones **??** y 5.2.3) se cumplen en el plano esférico (o proyectivo rígido); la diferencia esencial es que aquí las rectas se intersectan en un único punto y se pierde la redundancia que siempre acarreaban las parejas antípodas en $\mathbb{S}^2$. No vale la pena reescribirlos.

EJERCICIO 6.57 Argumenta por qué la definición (6.15) corresponde a la descripción previa de distancia en $\mathbb{P}^2$.

EJERCICIO 6.58 ¿Cuántos triángulos definen tres rectas en $\mathbb{P}^2$?

EJERCICIO 6.59 ¿Cuál es la fórmula del área de los triángulos en $\mathbb{P}^2$?

EJERCICIO 6.60 Da un ejemplo de tres puntos en $\mathbb{P}^2$ tales que las tres rectas que pasan por ellos forman sólo triángulos iguales. ¿En qué ángulos se intersectan estas rectas?

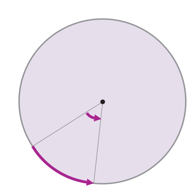
6.7.1 Las isometrías de $\mathbb{P}^2$ ($\mathbf{SO}(3)$)

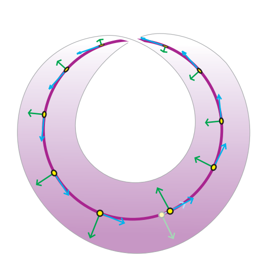
Podemos diferenciar al plano proyectivo holgado del rígido de manera "kleiniana", por el grupo de transformaciones que implícitamente se están considerando. Mientras que en el primero está uno pensando en el grupo de todas las proyectividades ($\mathbf{Pr}(2)$), en el segundo sólo se toma en cuenta al subgrupo de transformaciones proyectivas que preservan la métrica, es decir, a las isometrías.

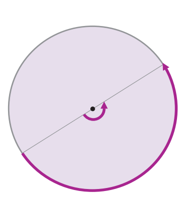
Puesto que la métrica de $\mathbb{P}^2$ se derivó de la de $\mathbb{S}^2$, entonces las transformaciones lineales que fijan a $\mathbb{S}^2$ y preservan su métrica (es decir, las ortogonales $\mathbf{O}(3)$) son las que dan lugar a isometrías de $\mathbb{P}^2$.

Notemos que la función antípoda $\mathfrak{a} \in \mathbf{O}(3)$ da lugar a la identidad en $\mathbb{P}^2$ y es la única que lo hace además de la identidad (Ejercicio 6.62). Entonces la proyectivización de f y de $\mathfrak{a} \circ \mathsf{f}$ son iguales para cualquier $\mathsf{f} \in \mathbf{O}(3)$, y sólo una de ellas está en $\mathbf{SO}(3)$ (preserva orientación de $\mathbb{S}^2$ y $\mathbb{R}^3$) pues la función antípoda $\mathfrak{a}$ invierte la orientación de $\mathbb{S}^2$. Por lo tanto, las isometrías de $\mathbb{P}^2$ están en correspondencia biunívoca con $\mathbf{SO}(3)$: los movimientos orientados de $\mathbb{S}^2$.

Esto implica (por el Teorema 5.14) que todas las isometrías de $\mathbb{P}^2$ son rotaciones. Efectivamente, al ir rotando en un punto, $[\mathbf{u}]$ digamos, su recta polar se va transladando la distancia correspondiente al ángulo (como en la esfera). Pero al llegar a la rotación de π en $[\mathbf{u}]$ su recta polar $\ell_{[\mathbf{u}]}$ ya regresó a su lugar punto a punto. En ese momento tenemos la reflexión en $\ell_{[\mathbf{u}]}$ (su espejo); localmente, en una vecindad de la recta, equivale a empujar una banda de Moebius dentro de sí misma una vuelta: el corazón regresa a su lugar pero los "dos" lados (tanto de la banda como del corazón y vistos localmente) se han intercambiado. Y a partir de π, seguir rotando en $[\mathbf{u}]$ parece un paso en $\ell_{[\mathbf{u}]}$ hasta dar una segunda vuelta y llegar a la identidad (rotar 2π). Así, los cuatro tipos de isometrías que teníamos en el plano euclidiano (y que, con algo más, reaparecerán en el plano hiperbólico) se vuelven lo mismo en el plano proyectivo rígido. En él sólo hay un tipo de isometrías, aunque se pueden ver desde distintos puntos de vista.







Notemos del párrafo anterior que hay tantas reflexiones en $\mathbb{P}^2$ como rectas; es decir, hay justo "un $\mathbb{P}^2$ de reflexiones" en $\mathbf{SO}\,(3)$. Y que muy cerca de cada reflexión están los pasos chicos en ambas posibles direcciones de su recta; este $\mathbb{P}^2$ de reflexiones se va "engordando" mientras consideramos pasos hasta un cierto tamaño creciente. Nótese tambien que los pasos de una longitud fija forman una esfera $\mathbb{S}^2$, pues corresponden a las rectas orientadas y que todos los pasos de longitud π son la identidad. Esta descripción remite a la de $\mathbb{P}^2$ a partir —alejándose— de un $\mathbb{P}^1$. Pero por el momento sólo nos interesa establecer que $\mathbf{SO}\,(3)$ se puede pensar como un espacio —hay una noción natural de cercanía en él (también en los otros grupos de transformaciones que hemos considerado)— y que este espacio se siente cerca de entender mejor...

EJERCICIO 6.61 Demuestra directamente que toda isometría de $\mathbb{P}^2$ es una rotación siguiendo las ideas de la demostración del Teorema 5.14. Verás que algunos argumentos se simplifican y que la demostración se limpia.

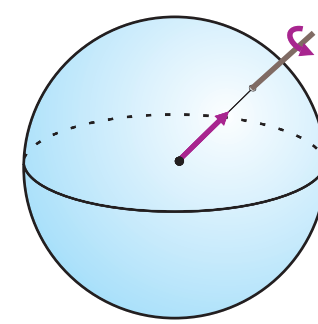
EJERCICIO 6.62 Con base en el Teorema de 4 en 4, demuestra que si la proyectivización de una transformación en $\mathbf{O}\,(3)$ es la identidad, entonces es la identidad o la función antípoda.

6.7.2 *El homeomorfismo de $\mathbf{SO}(3)$ con $\mathbb{P}^3$

Consideremos de nuevo la acción natural de $\mathbf{SO}(3)$ en la esfera $\mathbb{S}^2$ (sus movimientos rígidos). Recordemos de la Sección 5.3.1 que cada elemento de $\mathbf{SO}(3)$ es la rotación de un cierto ángulo α con eje en un vector unitario $\mathbf{u} \in \mathbb{S}^2$, donde α se mide de acuerdo con la orientación positiva definida en la superficie de la esfera. Además, a

tal rotación la habíamos denotado $\rho_{\alpha,\mathbf{u}}$ y se tiene que $\rho_{\alpha,\mathbf{u}} = \rho_{-\alpha,-\mathbf{u}}$, esto es, que al rotar positivamente en un punto de la esfera, se está rotando negativamente en su antípoda —es por esto que al intentar abrir un frasco de mermelada no importa con qué mano se tome la tapa, el movimiento (o el pujido) va para el mismo lado—.

Con esta notación, hemos coordinatizado con las "coordenadas polares" de $\mathbb{R}^3$ a los elementos del grupo $\mathbf{SO}(3)$. Un punto $\mathbf{x}$ en $\mathbb{R}^3$, tiene como *coordenadas polares* a un número no negativo α (con $\alpha = |\mathbf{x}|$) y a un vector unitario $\mathbf{u} \in \mathbb{S}^2$ ($\mathbf{u} = \mathbf{x}/|\mathbf{x}|$ cuando $\mathbf{x} \neq \mathbf{0}$ y puede ser cualquiera para $\mathbf{x} = \mathbf{0}$), de tal manera que $\mathbf{x} = \alpha\mathbf{u}$. Entonces, nuestra notación es una función natural



$$\rho : \mathbb{R}^3 \to \mathbf{SO}(3),$$

que en coordenadas polares se expresa como

$$\alpha, \mathbf{u} \mapsto \rho_{\alpha,\mathbf{u}}$$

y en coordenadas cartesianas tiene la fórmula

$$\mathbf{x} \mapsto \begin{cases} \mathrm{id}_{\mathbb{S}^2} & \text{si} \quad \mathbf{x} = \mathbf{0} \\ \rho_{|\mathbf{x}|,\, \mathbf{x}/|\mathbf{x}|} & \text{si} \quad \mathbf{x} \neq \mathbf{0}. \end{cases}$$

Nótese que esta función es "continua" en el origen, pues rotar cero grados en cualquier eje es la identidad y rotar con un ángulo muy pequeño está cercano a ella.

Esta función es "periódica". En la esfera de radio 2π es la constante identidad y a partir de ahí "vuelve a empezar". Pero incluso ahí ya tenemos "redundancia": ya le hemos "dado dos vueltas a $\mathbf{SO}(3)$", pues si llamamos B_π a la bola sólida con centro en el origen y radio π, es decir,

$$B_\pi = \left\{\mathbf{x} \in \mathbb{R}^3 \mid |\mathbf{x}| \leq \pi\right\},$$

entonces $\rho(B_\pi) = \mathbf{SO}(3)$. Con la bola sólida B_π ya cubrimos todas las rotaciones, pues una rotación de ángulo mayor que π se expresa en la antípoda con uno menor. Así que

$$\rho : B_\pi \to \mathbf{SO}(3)$$

es sobre y además es 1 a 1 excepto en la frontera, donde

$$\rho_{\pi,\mathbf{u}} = \rho_{\pi,-\mathbf{u}};$$

pues rotar π en una dirección es igual a rotar π en la contraria; nótese que corresponden a las reflexiones de $\mathbb{P}^2$ que vimos en la sección anterior.

Así que $\mathbf{SO}(3)$ *se puede pensar como la bola sólida* B_π *identificando puntos antípodas en la frontera*, pues estos dan lugar a la misma rotación (por un ángulo π). Por la analogía de que $\mathbb{P}^2$ se obtiene topológicamente de un disco identificando puntos antípodas en la frontera, este espacio debe ser el espacio proyectivo $\mathbb{P}^3$.

En algunos ejercicios hemos dejado al lector el problema de generalizar a 3 y más dimensiones los principios y las definiciones básicas de la geometría proyectiva; por cuestión de completez, en los párrafos que siguen esbozamos las ideas fundamentales.

Rudimentos de $\mathbb{P}^3$

Se define $\mathbb{P}^3$ como el espacio cuyos puntos corresponden a la líneas que pasan por el origen en $\mathbb{R}^4$. Si denotamos con $[\mathsf{x}]$ a la línea por el origen determinada por $\mathsf{x} \in \mathbb{R}^4$, con $\mathsf{x} \neq \mathsf{0}$ ($[\mathsf{x}] := \{t\,\mathsf{x} \mid t \in \mathbb{R}\}$), entonces

$$\mathbb{P}^3 := \left\{ \, [\mathsf{x}] \mid \mathsf{x} \in \mathbb{R}^4 \quad \text{con} \quad \mathsf{x} \neq \mathsf{0} \right\}$$

y tiene nociones naturales de línea (correspondiendo a planos por el origen en $\mathbb{R}^4$) y plano (las líneas en subespacios de dimensión 3). Se tiene además que cualquier par de planos se intersectan en una línea y una línea y un plano (que no la contenga) se intersectan siempre en un punto.

El espacio proyectivo se puede considerar con dos grupos naturales de transformaciones: el "holgado" que es con la *proyectivización* de todas las transformaciones lineales de $\mathbb{R}^4$ (es decir, considerando su acción en las líneas por el origen), y el "rígido" donde sólo se considera la proyectivización de las transformaciones ortogonales $\mathbf{O}(4)$ que preservan el producto interior y, por tanto, los ángulos que dan la distancia natural de $\mathbb{P}^3$. En el espacio proyectivo rígido, todos los planos son copias de $\mathbb{P}^2$ rígido y las líneas o geodésicas miden π.

Por su parte, también tenemos la esfera unitaria de dimensión 3 que consta de todos los vectores unitarios en $\mathbb{R}^4$,

$$\mathbb{S}^3 := \left\{ \, \mathsf{x} \in \mathbb{R}^4 \mid |\mathsf{x}| = 1 \, \right\},$$

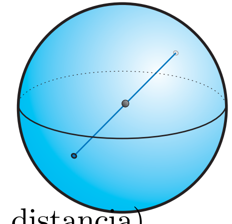
y es tal que cada recta por el origen en $\mathbb{R}^4$ la intersecta en dos puntos antípodas. Así que de nuevo hay un cubrimiento natural 2 a 1 de $\mathbb{S}^3$ sobre $\mathbb{P}^3$. Un *casquete* o *hemisferio* de $\mathbb{S}^3$ (el subconjunto donde alguna de sus coordenadas es no negativa) se puede identificar biyectivamente con una bola sólida en $\mathbb{R}^3$, y entonces se puede ver que el espacio proyectivo $\mathbb{P}^3$ es (topológicamente) una bola sólida con puntos antípodas en la frontera identificados. Esto es lo que necesitábamos para concluir que $\mathbf{SO}(3)$ *es topológicamente* $\mathbb{P}^3$ (se dice que son *homeomorfos,* "con la misma forma"), pero vale la pena hacer la identificación con más precisión.

Geometría de $\mathbf{SO}(3)$.

Puesto que $\mathbf{SO}(3)$ tiene estructura de grupo, veremos que tiene una "geometría" natural y que, al hacer las identificaciones con cuidado, ésta corresponde a la de $\mathbb{S}^3$.

Notemos primero que $\mathbf{SO}(3)$ tiene muchos subgrupos "circulares": dado $\mathbf{u} \in \mathbb{S}^2$, todas las rotaciones con eje en $\mathbf{u}$ (que coinciden con las de su antípoda $-\mathbf{u}$),

$$S_{\mathbf{u}} := \{\rho_{\alpha,\mathbf{u}}\},$$



forman un subgrupo que se comporta como el círculo $\mathbb{S}^1$ con la suma de ángulos (isomorfo a $\mathbf{SO}(2)$). Estos subgrupos son precisamente los diámetros al pensar a $\mathbf{SO}(3)$ como la bola B_π con identificación de antípodas en la frontera. En estos círculos hay una distancia natural, la ángular; y a partir de ella podremos extrapolar a una métrica (noción de distancia) en todo $\mathbf{SO}(3)$ haciéndolos geodésicos.

Como cada elemento de $\mathbf{SO}(3)$ está en uno de estos círculos (único si no es la identidad), entonces su distancia a $\mathrm{id}_{\mathbb{S}^2}$ debe ser el ángulo de rotación, o proporcional a él. Para que la distancia corresponda a la natural de $\mathbb{P}^3$, hay que tomar la mitad de los ángulos y definir, para $0 \leq \alpha \leq \pi$, que

$$d(\rho_{\alpha,\mathbf{u}}, \mathrm{id}) = \alpha/2. \tag{6.16}$$

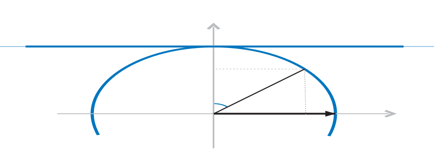
De aquí, aprovechando la estructura de grupo que nos da las transformaciones "multiplicar por un elemento constante", esparcimos la distancia definiendo para cualquier par $f, g \in \mathbf{SO}(3)$:

$$d(f, g) = d\left(f \circ g^{-1}, \mathrm{id}\right). \tag{6.17}$$

Pues al multiplicar un subgrupo circular (el que pasa por $f \circ g^{-1}$) por un elemento que no esté en él (g) nos da nuevas geodésicas (una que pasa por f y g).

Ahora sí, de las muchas maneras de identificar una bola sólida (B_π) con un casquete de $\mathbb{S}^3$ (por ejemplo la proyección estereográfica, proyecciones desde otros puntos o la ortogonal), la que geométricamente nos conviene es "*untarla radialmente*": mandar a los círculos $S_{\mathbf{u}}$ isométricamente. Esto es, despues de reescalar a B_π a la mitad, mandar a los diámetros por longitud de arco, que en coordenadas polares se escribe

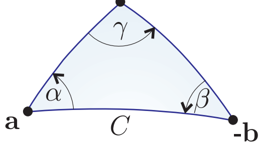
$$\begin{aligned} \mu : B_\pi &\to \mathbb{S}^3 \\ \mu(\alpha, \mathbf{u}) &= (\cos(\alpha/2), \mathrm{sen}(\alpha/2)\,\mathbf{u}). \end{aligned}$$



Donde estamos pensando a $\mathbb{R}^4$ como una pareja ordenada de un número real y un vector de $\mathbb{R}^3$ (la identificación natural $\mathbb{R}^4 = \mathbb{R} \times \mathbb{R}^3$); o bien añadirle una "coordenada cero" a un vector de $\mathbb{R}^3$, y estamos dibujando a esta "coordenada cero" verticalmente.

Veremos ahora que esta identificación μ *induce una isometría de* $\mathbf{SO}(3)$ *con* $\mathbb{P}^3$. Para esto, supongamos que f y g en $\mathbf{SO}(3)$ son

$$f = \rho_{2\alpha,\mathbf{a}} \quad \text{y} \quad g = \rho_{2\beta,\mathbf{b}}, \quad \text{con } \mathbf{a}, \mathbf{b} \in \mathbb{S}^2 \text{ y } 0 < \alpha, \beta \leq \pi/2.$$



Por (6.17) y (6.16), debemos medir el ángulo de $f \circ g^{-1}$. Puesto que g^{-1} es la rotación de 2β en el antípoda $-\mathbf{b}$, por el Lema 5.15 tenemos que $f \circ g^{-1} = \rho_{-2\gamma,\mathbf{c}} = \rho_{2\gamma,-\mathbf{c}}$, donde γ y $\mathbf{c}$ son el tercer ángulo y vértice del triángulo con vértices $\mathbf{a}$ y $-\mathbf{b}$ y ángulos correspondientes α y β. Sea $C = d_{\mathbb{S}^2}(\mathbf{a}, -\mathbf{b})$ lo que mide el lado opuesto a γ; por la Ley Esférica de Cosenos (Lema **??**) se tiene que

$$\cos\gamma = \cos C \ \text{sen}\,\alpha \ \text{sen}\,\beta - \cos\alpha \ \cos\beta.$$

Por el otro lado, la distancia en $\mathbb{S}^3$, que es la angular, se mide con el producto punto y se tiene

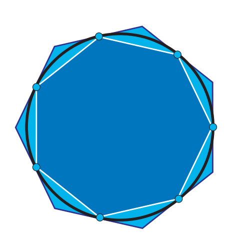
$$\begin{aligned}
\mu(2\alpha, \mathbf{a}) \cdot \mu(2\beta, \mathbf{b}) &= (\cos\alpha, \text{sen}\,\alpha \ \mathbf{a}) \cdot (\cos\beta, \text{sen}\,\beta \ \mathbf{b}) \\
&= \cos\alpha \ \cos\beta + \text{sen}\,\alpha \ \text{sen}\,\beta \ (\mathbf{a} \cdot \mathbf{b}) \\
&= \cos\alpha \ \cos\beta - \text{sen}\,\alpha \ \text{sen}\,\beta \ (\mathbf{a} \cdot (-\mathbf{b})) \\
&= \cos\alpha \ \cos\beta - \text{sen}\,\alpha \ \text{sen}\,\beta \ \cos C = \cos(\pi - \gamma) \,.
\end{aligned}$$

Tenemos entonces que γ (la mitad del ángulo de rotación de $f \circ g^{-1}$ medido en el triángulo) y el ángulo entre los vectores $\mu(2\alpha, \mathbf{a})$ y $\mu(2\beta, \mathbf{b})$ en $\mathbb{S}^3$ son complementarios. Pero en ambos casos, el menor de ellos es el que mide la distancia deseada: si $\gamma > \pi/2$ entonces $d(f, g) = \pi - \gamma$ (hay que considerarlo como rotación en la antípoda), y la distancia en $\mathbb{P}^3$ es el ángulo entre las rectas $[\mu(2\alpha, \mathbf{a})]$ y $[\mu(2\beta, \mathbf{b})]$ que es $\leq \pi/2$. Así que $\mathbf{SO}(3)$ y $\mathbb{P}^3$ son métricamente equivalentes o, dicho de otra manera, $\mathbb{P}^3$ tiene estructura de grupo compatible con su métrica natural.

Resulta que también $\mathbb{S}^3$ tiene estructura de grupo que se "levanta" de la de $\mathbf{SO}(3)$; así como $\mathbf{SO}(2)$ se puede pensar como los complejos unitarios, $\mathbb{S}^3$ resulta ser los cuaterniones unitarios.[4]

6.7.3 *Los politopos regulares de $\mathbb{R}^4$

Una observación interesante, debida a Coxeter en [5], que no podemos dejar pasar pues ya se hizo casi todo el trabajo, es que los subgrupos finitos de $\mathbf{SO}(3)$ dan lugar a subconjuntos finitos muy homogéneos en $\mathbb{S}^3$ al tomar su doble cubierta, y que de ellos se pueden construir los análogos de los sólidos platónicos en dimensión 4. Antes de verlo, considérese el ejemplo en dimensión 2 de n puntos equidistribuidos en $\mathbb{S}^1$ (un subgrupo finito de $\mathbf{SO}(2)$); al tomar la intersección de sus hemiplanos tangentes o bien su *casco convexo* (*i.e.*, lo que "atrapan" entre ellos o el conjunto de sus combinaciones afines positivas), se obtienen polígonos regulares de n lados: respectivamente *circunscrito* o *inscrito* en $\mathbb{S}^1$. Es fácil ver que los sólidos platónicos se pueden definir análogamente con subconjuntos finitos de $\mathbb{R}^3$; tienen esferas inscrita y circunscrita.



Ahora sí. Del grupo de simetrías orientadas del dodecaedro (que habíamos llamado 235 en la Sección 5.4.1) con 60 elementos se obtiene, al tomar su preimagen bajo la proyección $\mathbb{S}^3 \to \mathbb{P}^3 = \mathbf{SO}(3)$, un conjunto de 120 puntos en $\mathbb{S}^3 \subset \mathbb{R}^4$ con mucha simetría. Tan es así, que al tomar sus hiperplanos (de dimensión 3) tangentes a la esfera e intersectar sus hemiespacios, se obtiene uno de los "politopos regulares"

[4] Para seguir este camino ver [3].

que generalizan a los sólidos platónicos. Está formado por 120 dodecaedros (que habíamos denotado con $\{\{5,3\}\}$ en la Sección 5.4.3) pegados por sus caras y con 3 de ellos alrededor de cada arista; su símbolo de Schlafli es entonces $\{\{5,3,3\}\}$. Su dual se obtiene tomando el *casco convexo* de los 120 puntos: consta de 600 tetraedros con cinco alrededor de cada arista; es el $\{\{3,3,5\}\}$. En la literatura en inglés se les conoce como el "120-*cell*" y el "600-*cell*" respectivamente; Coxeter también les llama *panales* ("*honeycombs*"); en [2] se le llamó "*dodecaédromo*" al $\{\{5,3,3\}\}$.

Con el grupo de simetrías orientadas del tetraedro, el 233, y de cualquiera de las dos maneras (por tangentes o casco convexo), se obtiene el $\{\{3,4,3\}\}$ hecho de octaedros ($\{\{3,4\}\}$) y de a 3 por arista; resulta ser "autodual". El hipercubo $\{\{4,3,3\}\}$ se obtiene por tangentes a partir del grupo 222 con las rotaciones de π en los tres ejes; y su dual $\{\{3,3,4\}\}$ es el casco convexo.

Falta un politopo regular, el $\{\{3,3,3\}\}$ análogo del triángulo y el tetraedro con 5 vértices equidistantes. Y sobra un grupo importante: el del cubo. No todos, poliedros y grupos finitos, caen en este esquema... pero aquí lo dejamos, pues habíamos prometido no meternos demasiado en la dimensión 4.

6.8 *Despliegue de realidad virtual

La "realidad virtual" —esos juegos donde nos paseamos por un mundo imaginario, que está ahí, transcurriendo ante nuestros ojos en la pantalla de una computadora— consta, al menos, de tres grandes elementos. Primero, el escenario arquitectónico, la base de datos que lo define como una serie de objetos en tres dimensiones con sus posiciones, cualidades, texturas, etc. Segundo, que puede ser vacío, las reglas internas y los procesos que lo hacen cambiante y que permiten interactuar tanto a sus seres virtuales como al jugador. Y tercero, el procedimiento por el cual las porciones relevantes de estos datos se despliegan coherentemente en la pantalla para hacernos sentir que nos estamos moviendo, actuando y viendo lo que ocurre ahí. El problema que nos ocupa es este último, el de "ver" en una realidad virtual: desplegarla en la pantalla de una computadora; y, como veremos, es básicamente el problema del pintor con el que empezamos el capítulo pero hecho númerico y continuo. Además, su planteamiento teórico se aplica a otras muchas situaciones, como lo es parte del mecanismo básico de la visión humana.

La cámara

...Dibujo

Suponiendo que tenemos un "mundo virtual" definido en $\mathbb{R}^3$, ¿qué ve un punto que se pasea por él? Primero hay que definir que es eso de "*ver*": pensemos que deambulamos en ese mundo metidos dentro de una de esas cortinas negras de las cámaras daguerrotípicas de hace dos siglos, pero con una ventanita al frente (la pantalla de la computadora). *Ver* es asomarse el mundo desde esa ventana, o, lo que es lo mismo, proyectar al mundo externo en ella (la pantalla) que es parte

de un plano; y esta proyección se hace desde un punto (el ojo hipotético del sujeto ante la pantalla, o debajo de la cortina): es, entonces, el problema del pintor con que arrancamos el capítulo. Pero tenemos aún más; ese plano está coordinatizado. La pantalla de una computadora consta de pixeles determinados por dos coordenadas y a ellos habrá que asignárseles un color. Nótese, además, que también un lienzo en blanco, la película (o el detector digital) en una cámara o bien la retina de un ojo (suponiéndola plana para incluirnos en la discusión) tienen sus coordenadas locales establecidas. Entonces hay que pensar que en ese plano hay un origen y dos vectores direccionales distinguidos, $\mathbf{u}_1$ y $\mathbf{u}_2$ digamos, que lo generan y coordinatizan.

Además, está el punto desde el cual se proyecta, llamémoslo el *foco,* que tiene una posición bien definida $\mathbf{u}_4 \in \mathbb{R}^3$; y para acabar de definir el plano y la proyección, tomemos a $\mathbf{u}_3$ como el vector que va del foco al origen del plano pantalla. Estos cuatro vectores definen una "*cámara*".

...Dibujo

En la vida real (cámaras, telescopios, ojos, etc.) $\mathbf{u}_1$ y $\mathbf{u}_2$ son de tamaño fijo. En un ojo, $\mathbf{u}_3$ también lo es y apunta hacia atrás del foco (la retina está detrás del foco que es simulado por un sistema óptico). En una cámara de cine —donde incluimos la fotográfica— $\mathbf{u}_3$ también apunta hacia atrás pero es variable: su tamaño es la *distancia focal*. Al crecer se tiene el efecto de *zoom*, se hace telefoto; y al decrecer tiende al gran angular pues más de la escena entra a la parte sensible de la pantalla (al acercarse a una ventana aumenta la porción de mundo que se ve en ella).

Por su parte, tanto en la computadora como en el caso del pintor, $\mathbf{u}_3$ apunta hacia adelante, hacia la escena relevante. En la computadora, $\mathbf{u}_1$ y $\mathbf{u}_2$ son las coordenadas de la pantalla y $\mathbf{u}_3$ representa la distancia del usuario promedio a ella; aunque al hacerla crecer se logra el efecto de acercamiento o magnificación (sin que el usuario se tenga que mover de su lugar). Este efecto también se puede lograr en la computadora disminuyendo los vectores $\mathbf{u}_1$ y $\mathbf{u}_2$ que definen las coordenadas de la pantalla (y seguir pensando que el usuario sigue en su silla). Nótese, por último, que cambiar $\mathbf{u}_4$ corresponde justamente a desplazar la cámara, paralela a sí misma, dentro del mundo; da la posibilidad de "pasearse" por él. Y aquí sí, ni modo, se trivializa para el pintor y su caballete fijo.

...Dibujo

En la *cámara canónica*, cuando $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ es la base canónica y $\mathbf{u}_4$ es el origen, conviene pensar a las dos primeras coordenadas, la x y la y, como las de un plano vertical y a la coordenada z como la profundidad (las z positivas son las que interesan cuando pensamos en el lienzo o en la pantalla de la computadora, y las negativas para la película o la retina). Entonces la función *proyección* es muy sencilla de expresar:

$$\begin{aligned} \text{“}\ \varphi &: \mathbb{R}^3 \to \mathbb{R}^2\ \text{”} \\ \varphi(x, y, z) &= (x/z, y/z)\,. \end{aligned}$$

Pues la intersección de la recta por el origen (el foco) y el punto "real" (x, y, z) con la pantalla (que es el plano $z = 1$) es $(x/z, y/z, 1)$, pero de éste sus primeras dos

coordenadas son las de la pantalla. El problema general se resuelve a partir de éste.

Dada una cámara cualquiera $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4$ y un punto $\mathbf{x} \in \mathbb{R}^3$, su imagen se obtiene aplicando φ a las coordenadas con que se expresa $\mathbf{x}$ en las coordenadas afines del sistema de referencia dado por la cámara. Y el problema númerico se reduce a encontrar tales coordenadas; hay que expresar el mundo en ese sistema y dividir a las dos primeras coordenadas entre la tercera.

6.8.1 Proyecciones del espacio proyectivo

Aunque en principio hemos resuelto el problema, veamos que al proyectivizar éste adquiere más consistencia y la solución numérica se hace más fácil en la práctica.

Primero tenemos que pensar en $\mathbb{R}^3$ como la carta canónica de $\mathbb{P}^3$, que analíticamente es añadir un 1 a las tres coordenadas y pensar a estas como homogéneas. Geométricamente, se añade entonces un plano proyectivo al infinito con coordenadas homogéneas $[x : y : z : 0]$, que, si llamamos w a la nueva coordenada homogénea, está determinado por la ecuación $w = 0$. Un punto $[x : y : z : w] \in \mathbb{P}^3$ que no tenga su última coordenada cero representa al punto $(x/w, y/w, z/w) \in \mathbb{R}^3$.

En términos de definir una realidad virtual, pensar en $\mathbb{R}^3$ como parte del espacio proyectivo añadiéndole un plano proyectivo al infinito es importante, pues permite agregar un *horizonte*; un fondo en el infinito, una textura fija definida por las posibles direcciones. Parecería que hay una contradicción con que en el espacio proyectivo $\mathbb{P}^3$ se llega al mismo punto al infinito viajando en direcciones opuestas. Pero no hay tal contradicción. En efecto es el mismo punto, pero cada plano tiene localmente dos lados y a estos puede asignárseles colores diferentes.

La información de una cámara equivale a la matriz de 4×4

$$C = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

donde los datos reales son la matriz superior de 3×4 y a la cual se añade un renglón con 0's y un 1 al final. Análogamente a la Sección 6.6.4, esta matriz representa en $\mathbb{P}^3$ a la transformación afín de $\mathbb{R}^3$ que manda a la cámara canónica en la dada, pues obsérvese que manda al vector $(x, y, z, 1)$ en

$$x\,\mathbf{u}_1 + y\,\mathbf{u}_2 + z\,\mathbf{u}_3 + \mathbf{u}_4.$$

Lo que se necesita entonces, para dar númericamente a la proyección determinada por la cámara C, es encontrar su matriz inversa y luego aplicar la proyección canónica; es decir, se tendrá la fórmula

$$\varphi\left(C^{-1}\mathbf{x}\right).$$

Parecería muy difícil sacar la matriz inversa de una de 4×4, pero en la práctica se puede hacer en forma muy sencilla. Nótese primero que la identidad I corresponde a la cámara canónica que es donde, para simplificar, se puede empezar un recorrido

por el mundo virtual. Y segundo, que los movimientos que producen las teclas correponden a matrices muy sencillas. Por ejemplo, las flechas izquierda y derecha, o los correspondientes movimientos del ratón, son rotaciones en el plano $\mathbf{u}_1, \mathbf{u}_3$ (voltear hacia arriba o hacia abajo son rotaciones en el plano $\mathbf{u}_2, \mathbf{u}_3$) cuyas matrices son muy simples y sus inversas se obtienen casi trivialmente. Así que si tenemos una cámara dada C, un teclazo de estos cambiará a C^{-1} multiplicándola (por el lado adecuado: ¿cuál es éste?) por la correspondiente matriz de rotación. De la misma manera, moverse hacía adelante es cambiar $\mathbf{u}_4$ por $\mathbf{u}_4 + t\,\mathbf{u}_3$, cuya matriz correspondiente y su inversa son muy sencillas. Entonces, al ir apretando las teclas, los coeficientes de la matriz importante para la proyección, que es C^{-1}, van cambiando poco a poco y de manera fácil de obtener.

Capítulo 7

Cónicas III (proyectivas)

En este capítulo retomamos la motivación que surgió de las cónicas en $\mathbb{R}^2$ para estudiar el "infinito" y que nos llevó a definir la geometría proyectiva. Ya tenemos la herramienta para entender y formalizar el comportamiento de las cónicas en el infinito, pero con el mismo esfuerzo se puede hacer una buena parte de la teoría para curvas algebraicas en general, y así lo haremos. Después clasificaremos las cónicas proyectivamente y esbozaremos qué debe pasar en dimensiones más altas.

7.1 Curvas algebraicas en $\mathbb{P}^2$

Lo que nos interesa principalmente es la completación proyectiva de las curvas cuadráticas, es decir, entender el comportamiento de las cónicas al infinito, pero hay ciertas partes de la teoría que se pueden generalizar fácilmente, y vale la pena hacerlo para no repetir *verbatim* lo que vimos en el Capítulo 4. Ahí definimos una curva cuadrática como los ceros de un polinomio cuadrático en dos variables x y y. Es claro que esto puede generalizarse a cualquier polinomio y eso haremos.

7.1.1 Polinomios y su homogeneización

Un polinomio en una variable x es una expresión

$$P(x) = \sum_{i=0}^{k} a_i x^i.$$

Su *grado* es la máxima i para la cual $a_i \neq 0$, y cada sumando es un *monomio*. Un *monomio* en dos variables x y y es una expresión de la forma

$$a_{i,j}\, x^i y^j,$$

donde a_{ij} es un número real; su *grado* es $i+j$; y un polinomio en dos variables x y y es una suma finita de monomios. Así que se puede expresar como una sumatoria

$$P(x,y) = \sum_{i,j=0}^{k} a_{i,j}\, x^i y^j, \tag{7.1}$$

donde quizá muchos de los *coeficientes* a_{ij} son 0. El *grado* de un polinomio en dos variables es el máximo grado de sus monomios (que no necesariamente coincide con el límite superior k de la sumatoria anterior, que se puso ahí sólo para enfatizar que hay un número finito de sumandos).

Los polinomios en dos variables (como los de una) tienen una estructura algebraica muy rica. Se pueden sumar (y con la suma forman un grupo) y se pueden multiplicar (aunque no forman un grupo multiplicativo pues no tienen inversos). En este sentido, son como los números enteros, $\mathbb{Z}$, su unidad multiplicativa es el polinomio constante 1, y por esto se le conoce como el *anillo* de los polinomios.

Vale la pena declarar de una vez por todas que cuando digamos "polinomio" sin especificar el apellido "constante" nos referimos a un polinomio de "a de veras", de grado al menos 1. Los de grado 0 (las constantes distintas de cero) y el de grado -1 (la constante 0 a la que hay que ponerle este grado para que las cosas funcionen) no se entrometerán tanto en el texto con esta convención.

Como en el caso de los polinomios en una variable, los polinomios en dos variables se pueden considerar también como funciones con valores reales (al evaluar sustituyendo valores fijos de las variables), pero en este caso su dominio es $\mathbb{R}^2$. Una *curva algebraica* en $\mathbb{R}^2$ es el conjunto de ceros de un polinomio P, es decir

$$C(P) := \left\{(x,y) \in \mathbb{R}^2 \mid P(x,y) = 0\right\}.$$

El *grado* de una curva algebraica es el grado de su polinomio asociado. Ya hemos clasificado las curvas algebraicas de grados 1 y 2 (rectas y cónicas). El problema natural de entender y clasificar todas las curvas algebraicas ha dado lugar a lo que ahora se conoce como **geometría algebraica** que claramente requiere un estudio más profundo del álgebra subyacente. Por el momento, veamos una relación muy sencilla entre el álgebra y la geometría.

Supongamos que un polinomio P es el producto de dos polinomios P_1 y P_2. Es decir, que P se obtiene al multiplicar, con las reglas de los número reales, las expresiones de P_1 y de P_2 (todos los monomios de uno contra los del otro y luego agrupando). Esto se escribe

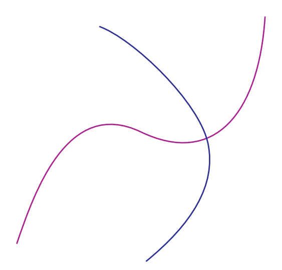
$$P(x,y) = P_1(x,y)\, P_2(x,y).$$

Entonces es cierto que para cualquier valor de las variables x, y esta igualdad se cumple. En particular se obtiene que

$$P(x,y) = 0 \quad \Leftrightarrow \quad P_1(x,y) = 0 \quad \text{o} \quad P_2(x,y) = 0,$$

y por lo tanto que

$$C(P) = C(P_1) \cup C(P_2).$$



Hemos demostrado que si P es *reducible* (es decir, si P es un producto de polinomios, de grado ≥ 1, recuérdese nuestra convención), entonces su curva algebraica asociada es unión de curvas algebraicas. Esto explica en general por qué cuando clasificamos curvas cuadráticas nos aparecían a veces parejas de rectas; surgen de la multiplicación de dos polinomios lineales. Es natural definir entonces que una curva algebraica es *irreducible* si su polinomio asociado lo es, es decir, si no se puede expresar como un producto de polinomios (de grado al menos 1, insistimos por última vez).

Es muy claro que también tenemos polinomios en tres (o n) variables; que tienen su grado, su suma y multiplicación y que sus ceros definen subconjuntos de $\mathbb{R}^3$ (o, respectivamente, de $\mathbb{R}^n$) a los que se les llama *superficies algebraicas* (o bien, *hipersuperficies algebraicas*). Por el momento, nos interesa definir un polinomio de tres variables a partir de uno de dos.

Dado P como en (7.1), su *homogeneización*, o bien su *polinomio homogéneo asociado,* es

$$\tilde{P}(x, y, z) = \sum_{i,j=0}^{k} a_{i,j}\, x^i y^j z^{(n-i-j)},$$

donde n es el grado de P.

El nombre viene de que, en general, se dice que un polinomio es *homogéneo* de grado n si todos sus monomios son de grado n. Y la receta para obtener el polinomio homogéneo asociado es muy fácil: cada monomio se multiplica por la potencia de z necesaria para convertirlo en monomio de grado n. De su polinomio homogéneo se puede recuperar el polinomio original al hacer $z = 1$; es decir, se tiene que

$$P(x, y) = \tilde{P}(x, y, 1). \tag{7.2}$$

De tal manera que la asociación $P \mapsto \tilde{P}$ es inyectiva de los polinomios de grado n en dos variables a los homogéneos de grado n en tres variables. Y un polinomio homogéneo en tres variables viene de uno en dos si y sólo si z no aparece en todos sus monomios, es decir, si y sólo si el polinomio z no lo divide. Esto se da pues a los monomios de grado n de P (que los hay) no se les multiplica por nada ($z^0 = 1$).

EJERCICIO 7.1 Argumenta geométricamente que los polinomios x^2+y^2-1, x^2-y^2-1 y x^2-y son irreducibles.

EJERCICIO 7.2 ¿Cuáles son los polinomios homogéneos asociados a los anteriores? ¿Puedes describir las superficies que definen en $\mathbb{R}^3$? Describe cuáles son sus intersecciones con los planos horizontales $z = \mathrm{cte}$; y luego qué pasa cuando z varía.

EJERCICIO 7.3 Demuestra que la superficie algebraica definida por el polinomio homogéneo asociado a un polinomio lineal en dos variables es un plano por el origen.

7.1.2 Conos y curvas algebraicas proyectivas

Veamos ahora cómo es la superficie algebraica definida en $\mathbb{R}^3$ por la homogeneización de un polinomio en dos variables.

Hay que explicitar primero lo que dijimos antes por ensimita. Dado cualquier polinomio en tres variables $P(x,y,z)$, define un subconjunto de $\mathbb{R}^3$,

$$C(P) = \left\{(x,y,z) \in \mathbb{R}^3 \mid P(x,y,z) = 0\right\},$$

llamado su *superficie asociada*. Y a estos subconjuntos se les conoce como las *superficies algebraicas* de $\mathbb{R}^3$. Un ejemplo que ya conocemos bien son los planos, ya que están definidos por polinomios lineales. Otro ejemplo que ya hemos estudiado es la esfera unitaria $\mathbb{S}^2 = \left\{(x,y,z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 - 1 = 0\right\}$; o bien, cualquier esfera, pues estará definida por un polinomio de la forma $(x-a)^2 + (y-b)^2 + (z-c)^2 - d^2$ (ver más ejemplos en los siguientes ejercicios).

Supongamos ahora que $\tilde{P}(x,y,z)$ **es un polinomio homogéneo en tres variables**. Se cumple entonces que $\mathbf{0} \in C(\tilde{P})$, pues como cada monomio tiene al menos una variable, se anulará en el origen; esto es, $\tilde{P}(0,0,0) = 0$. Pero además, la propiedad de ser homogéneo nos da que para cualquier $t \in \mathbb{R}$ se cumple que

$$\tilde{P}(tx, ty, tz) = t^n\, \tilde{P}(x,y,z), \tag{7.3}$$

donde n es el grado de $\tilde{P}$; pues en cada monomio se asocian las t correspondientes a las tres variables para darnos una t^n que se factoriza en toda la sumatoria. Esta propiedad es exclusiva de los polinomios homogéneos, pues en los otros las t de los monomios adquieren diferentes exponentes y es imposible factorizarlos más allá de los pedazos homogéneos de un grado dado.

Esta propiedad tiene una implicación geométrica importantísima:

$$C(\tilde{P})\, \textit{es unión de rectas por el origen.}$$

Pues (7.3) implica claramente que

$$\tilde{P}(x,y,z) = 0 \quad \Leftrightarrow \quad \tilde{P}(tx,ty,tz) = 0,\ \forall\, t \in \mathbb{R}.$$

Conviene introducir la notación de un vector variable $\mathbf{x} = (x,y,z)$ con la obvia $\tilde{P}\,(\mathbf{x}) = \tilde{P}(x,y,z)$, para poder concluir que si $\mathbf{x} \neq \mathbf{0}$ (nada más para estar de acuerdo con nuestra convención del capítulo anterior), entonces

$$\mathbf{x} \in C(\tilde{P}) \quad \Leftrightarrow \quad [\mathbf{x}] \subset C(\tilde{P}),$$

donde $[\mathbf{x}]$ está pensado como recta por el origen en $\mathbb{R}^3$. Pero si lo pensamos como punto en $\mathbb{P}^2$, nos lleva a la definición básica de este capítulo.

Definición 7.1.1 Una *curva algebraica* de grado n (≥ 1) en el plano proyectivo es un subconjunto definido como

$$\tilde{C}(\tilde{P}) = \left\{ [\mathbf{x}] \in \mathbb{P}^2 \mid \tilde{P}(\mathbf{x}) = 0 \right\},$$

donde $\tilde{P}(x, y, z) = \tilde{P}(\mathbf{x})$ es un polinomio homogéneo de grado n. Si además $\tilde{P}$ es la homogeneización de un polinomio $P(x, y)$, entonces a $\tilde{C}(\tilde{P})$ se le llama la *cerradura proyectiva* de la curva algebraica $C(P) \subset \mathbb{R}^2$.

Puesto que $C(\tilde{P}) \subset \mathbb{R}^3$ es unión de rectas por el origen, se le llama *cono*. Y esta misma razón nos permite pensarlo como un conjunto bien definido de puntos en $\mathbb{P}^2$, que denotamos $\tilde{C}(\tilde{P})$. En el caso que $\tilde{P}$ se obtiene de un polinomio en dos variables $P(x, y)$ como su homogeneización, entonces (como demostraremos a continuación) este cono consiste exactamente de las rectas que pasan por el origen y los puntos de la curva $C(P)$ dibujada en el plano coordenado canónico $z = 1$ (es el *cono sobre* $C(P)$); y apenas incluye unas cuantas rectas horizontales más que corresponden a cómo se aproxima la curva al infinito.

Teorema 7.1 *Sea* $P(x, y)$ *un polinomio de grado* n*. Entonces su curva algebraica y su cerradura proyectiva coinciden en* $\mathbb{R}^2$ *(pensado dentro de* $\mathbb{P}^2$ *como la carta coordenada canónica* $z = 1$*), es decir,*

$$C(P) = \tilde{C}(\tilde{P}) \cap \mathbb{R}^2$$

y además, la cerradura proyectiva tiene a lo sumo n *puntos en el infinito, es decir*

$$\sharp(\tilde{C}(\tilde{P}) - C(P)) \leq n.$$

Demostración. Sea Π_1 el plano $z = 1$ en $\mathbb{R}^3$, e identificamos sus puntos con los de $\mathbb{R}^2$ como siempre (mediante la correspondencia $(x, y, 1) \leftrightarrow (x, y)$).

Demostremos primero que

$$C(P) = C(\tilde{P}) \cap \Pi_1. \tag{7.4}$$

Puesto que $P(x, y) = \tilde{P}(x, y, 1)$ (recúerdese (7.2)), se tiene que

$$P(x, y) = 0 \quad \Leftrightarrow \quad \tilde{P}(x, y, 1) = 0.$$

La expresión de la izquierda dice que $(x, y) \in C(P)$ y la de la derecha que su punto correspondiente en Π_1, $(x, y, 1)$, está en $C(\tilde{P})$. Lo cual demuestra (7.4).

Consideremos ahora el primer enunciado del teorema. Si $[\mathbf{x}] \in \tilde{C}(\tilde{P}) \cap \mathbb{R}^2$, entonces está representado por un vector $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$ tal que $z \neq 0$ y que además está en el cono $C(\tilde{P})$. Pero entonces $(x/z, y/z, 1)$ también está en $C(\tilde{P})$, y por (7.4), está en la curva original $C(P)$; por tanto $\tilde{C}(\tilde{P}) \cap \mathbb{R}^2 \subset C(P)$. El otro lado es más fácil, si $(x, y) \in C(P)$, entonces $\tilde{P}(x, y, 1) = P(x, y) = 0$ y por definición $[x : y : 1] \in \tilde{C}(\tilde{P})$.

Entonces, lo único nuevo que puede tener $\tilde{C}(\tilde{P})$ (que no tenía $C(P)$) está en el infinito, es decir, puntos de la forma $[x : y : 0]$. Para estudiarlos, consideremos el plano Π_0 paralelo a Π_1 por el origen, dado por la ecuación $z = 0$. Nos interesa ahora describir el conjunto

$$C(\tilde{P}) \cap \Pi_0.$$

Sabemos que consiste de rectas por el origen. ¿Cuáles? Para saberlo, hay que sustituir $z = 0$ en el polinomio $\tilde{P}$: esto anula todos los monomios de P excepto los de grado n, que son los únicos que no multiplicamos por z (o una potencia mayor de z) para obtener $\tilde{P}$. Ésta es la parte *dominante o de grado máximo* del polinomio P, llamémosle P_n. Es decir,

$$P_n(x, y) := \tilde{P}(x, y, 0).$$

Podemos identificar a $C(\tilde{P}) \cap \Pi_0$ con $C(P_n)$, y ello por la correspondencia obvia $(x, y, 0) \leftrightarrow (x, y)$. Como P_n es homogéneo de grado n (aunque ahora de dos variables) se corrobora que $C(P_n)$ consiste de rectas por el origen en un plano. Nos falta ver que son a lo más n para completar el teorema.

Supongamos primero que el polinomio y no divide a P_n, es decir, que no todos los monomios de P_n tienen a y (o, lo que es lo mismo, que el coeficiente de x^n es distinto de cero). Esto implica que la recta $y = 0$ no está en $C(P_n)$ pues $P_n(1, 0) \neq 0$ (es justo el coeficiente de x^n). Se tiene entonces que todas las rectas de $C(P_n)$ intersectan a la recta $y = 1$ en un punto que además las determina. Estos puntos se encuentran al resolver la ecuación en x

$$P_n(x, 1) = 0.$$

Como $P_n(x, 1)$ es un polinomio de grado n en una variable, tiene a lo más n raíces reales, y hemos terminado.

Nos falta el caso en que y divide a P_n, pero es igual de fácil. Ya tenemos una recta de $C(P_n)$, pues $P_n(x, 0) = 0$. Al factorizar y nos queda un polinomio de grado menor que n y sus soluciones, como en el caso anterior, serán a lo más $n - 1$. □

Observación. Aunque hemos usado los polinomios homogéneos para definir las curvas algebraicas proyectivas, estos **no** definen funciones de $\mathbb{P}^2$; hay que resistir la tentación pues, a diferencia del caso euclidiano, no tienen valores bien definidos excepto sus ceros. Nótese que el valor de un polinomio homogéneo en tres variables cambia dentro de una línea por el origen en $\mathbb{R}^3$ esencialmente como lo hace t^n; ahí no es constante y por tanto no se le puede asignar un valor al punto proyectivo correspondiente excepto cuando es cero.

Ejemplos

Dibujo...

Líneas rectas Sabemos que cualquier línea recta de $\mathbb{R}^2$ está definida por un polinomio lineal $ax + by + c$. Su homogeneización $ax + by + cz$ define precisamente su

completación proyectiva que intersecta a la línea al infinito en el punto $[-b : a : 0]$. De tal manera que lo que hemos hecho incluye lo que ya sabíamos del Capítulo 6 sobre los polinomios lineales y las rectas.

Cónicas Sea

$$P(x, y) = x^2 + y^2 - 1.$$

Sabemos bien que $C(P)$ es el **círculo** unitario $\mathbb{S}^1$. Su polinomio homogéneo asociado es

$$\tilde{P}(x, y, z) = x^2 + y^2 - z^2,$$

que define el cono desde el origen del círculo unitario dibujado en el plano $z = 1$. ...Dibujo Esto se puede ver desde la óptica de la demostración anterior, o bien observando que al variar z se obtienen círculos centrados en el origen con radio $|z|$. La parte dominante de P es $P_2(x, y) = x^2 + y^2$, cuya única solución es $(0, 0)$, la que no vale proyectivamente. De tal manera que su completación proyectiva es ella misma, $\mathbb{S}^1$ no tiene nada al infinito; lo cual es bastante obvio, pero corrobora que nuestro método es bueno: no añade nada si no hay necesidad.

Consideremos ahora a la **hipérbola**; sea

$$P(x, y) = x^2 - y^2 - 1,$$

cuyo polinomio homogéneo asociado es

$$\tilde{P}(x, y, z) = x^2 - y^2 - z^2.$$

En el infinito, cuando $z = 0$, tenemos la parte dominante ...Dibujo

$$P_2(x, y) = x^2 - y^2 = (x + y)(x - y),$$

que define dos rectas, las asíntotas. De tal manera que a la completación proyectiva de la hipérbola, $\tilde{C}(\tilde{P})$, se le añaden en el infinito precisamente los puntos correspondientes a sus asíntotas. El cono $C(\tilde{P})$ es de nuevo un cono circular pero ahora horizontal que intersecta al plano $z = 0$ en las asíntotas. Para ver esto, obsérvese que su intersección con el plano $x = 1$ es un círculo unitario; o en general, que sus intersecciones con los planos $x = \mathrm{cte}$ son círculos que crecen linealmente (la coordenada x desempeña ahora el papel que tenía la coordenada z en el caso anterior). Pensando en el plano proyectivo, la hipérbola se ve como círculo en la carta coordenada $x = 1$. Proyectivamente es igual al círculo.

Veamos por último a la **parábola** definida por el polinomio

$$P(x, y) = x^2 - 2y, \tag{7.5}$$

cuyo polinomio homogéneo asociado es

$$\tilde{P}(x, y, z) = x^2 - 2zy. \tag{7.6}$$

Su parte dominante es $P_2(x, y) = x^2$, cuya única solución es la recta $x = 0$ (el eje y). Dibujo... Así que, como era de esperarse, la parábola intersecta al infinito en un solo punto: el correspondiente a su eje. En este caso, el cono $C(\tilde{P})$ vuelve a ser un cono circular, pero ahora "descansa" sobre el plano $z = 0$; es decir, lo toca únicamente en una línea. También contiene al eje z, y los planos donde se intersecta en un círculo tienen la ecuación $y + z = k$, con k constante (ver Ejercicio 7.4).

Obsérvese que en este caso el cono $C(\tilde{P})$ es tangente al plano $z = 0$, lo cual quiere decir que la línea al infinito es tangente a la cónica proyectiva $\tilde{C}(\tilde{P})$. Esto es, las parábolas son tangentes a la línea al infinito y la tocan en un punto único.

El Teorema de Bézout (descripción)

Ya con algunos ejemplos concretos en la cabeza, esbocemos algunas consecuencias del Teorema 7.1.

Primero, como ya habíamos visto, los únicos polinomios homogéneos que no son la homogeneización de un polinomio en dos variables son aquellos divisibles por z; y que entonces tienen a la recta al infinito $z = 0$ como componente. Pero la recta al infinito no tiene nada de especial dentro del plano proyectivo. Así que una curva algebráica proyectiva o bien contiene una recta dada (como componente), o bien la intersecta en un número finito de puntos y este número es a lo más el grado de la curva. Hemos visto ya tres ejemplos donde este número varía entre 0 y el grado de la curva. Si hay menos puntos de intersección que el grado es porque algún polinomio no tuvo todas sus raíces potenciales; algunas de ellas fueron complejas, y esto es porque los números reales no son "algebraicamente cerrados".

Los números complejos sí son algebraicamente cerrados, es decir, cualquier polinomio (en una variable) tiene tantas raíces como su grado.[1] En el siglo XIX se desarrolló su teoría a la par que explotaba la geometría. Las dos ideas se lograron juntar dando lugar a lo que ahora se conoce como geometría algebraica. En ella se siguen (y de ella surgen) las ideas de construir una geometría proyectiva con coordenadas homogéneas y básicamente las mismas definiciones que hemos dado, pero sobre el campo de los números complejos. Y en ese contexto, sí es cierto lo que se intuye, si dos curvas algebraicas no tienen componentes comunes, entonces se intersectan en tantos puntos (contados adecuadamente con multiplicidad) como el producto de sus grados. A esto se le conoce como el Teorema de Bézout, pero rebasa el nivel de este libro; que quede ahí como motivación.

EJERCICIO 7.4 Demuestra que la intersección del cono $C(\tilde{P})$ de la parábola, donde $\tilde{P}(x, y) = x^2 - 2y$, con el el plano $y + z = k$ (con k constante) es un círculo, y da una parametrización isométrica de él. ¿Cuál es el radio de ese círculo?

[1] Ver Apéndice.

EJERCICIO 7.5 Considera la familia de polinomios $P_k(x, y) = x^2 - y^2 - k$, donde k es una constante real. Demuestra que las curvas algebraicas de todos estos polinomios intersectan a la recta al infinito en los mismos dos puntos: ¿cuáles son? ¿Puedes dibujar qué tipos de curvas definen en $\mathbb{R}^2$ cuando k es negativa y cuando k es positiva? ¿Cuando es chica o grande?

EJERCICIO 7.6 Considera la familia de polinomios $P_k(x, y) = x^3 - xy^2 - k$, donde k es una constante real. Demuestra que las curvas algebraicas de todos estos polinomios intersectan a la recta al infinito en los mismos tres puntos: ¿cuáles son? ¿Puedes dibujar, a grandes rasgos, qué tipos de curvas definen en $\mathbb{R}^2$ cuando k es negativa y cuando k es positiva? Para esto, puedes encontrar los sectores de $\mathbb{R}^2$ donde el polinomio $P_0(x, y)$ es positivo o negativo.

7.1.3 Equivalencia proyectiva

Así como definimos la equivalencia de curvas cuadráticas en $\mathbb{R}^2$ usando el grupo afín, para las curvas algebraicas proyectivas tenemos que usar el grupo de transformaciones proyectivas. La equivalencia geométrica puede muy bien definirse en general para subconjuntos.

Dados dos subconjuntos C y C' del plano proyectivo, se dice que son *equivalentes*, y lo escribimos $C \sim C'$, si existe $f \in \mathbf{Pr}(2)$ tal que $f(C') = C$, *i.e,* si son $\mathbf{Pr}(2)$-equivalentes como figuras. Así, por ejemplo, todos los cuadrángulos (cuatro puntos en posición general) son equivalentes, o por dualidad, cuatro rectas sin tercias concurrentes lo son. Es fácil ver que ésta es una relación de equivalencia (ver Ejercicio 7.7).

Uno de los problemas que dan lugar al desarrollo de la geometría algebraica es la clasificación de curvas algebraicas, entendida como la descripción o enumeración de sus clases de equivalencia. Como la herramienta para definirlas o trabajarlas son los polinomios, es necesario definir la equivalencia de polinomios homogéneos.

Sean $P(x, y, z)$ y $Q(x, y, z)$ dos polinomios homogéneos. Se dice que son *proyectivamente equivalentes*, o simplemente *equivalentes*, si existen $t \neq 0$ y una transformación lineal $f : \mathbb{R}^3 \to \mathbb{R}^3$ tales que

$$Q = t\,(P \circ f).$$

Puesto que f se escribe en coordenadas como $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x}))$ donde cada f_i $(i = 1, 2, 3)$ es una función lineal de la forma $f_i(x, y, z) = a_i x + b_i y + c_i z$, entonces $P \circ f$ se obtiene sustituyendo estos polinomios lineales en vez de sus tres variables y claramente resulta un nuevo polinomio homogéneo y del mismo grado. Así que la noción de equivalencia está bien definida y sólo hace equivalentes a polinomios de un mismo grado.

Por ejemplo, si $P(x, y, z) = x^2 + y^2 - z^2$ y $Q(x, y, z) = x^2 - y^2 - z^2$ son los polinomios que vienen del círculo y la hipérbola respectivamente, entonces para ver

que son equivalentes tomemos $f(x,y,z) = (z,y,x)$ y $t = -1$, pues entonces

$$\begin{aligned} t\ (P \circ f)(x,y,z) &= -1\,(P(f(x,y,z))) \\ &= -P(z,y,x) = -\left(z^2 + y^2 - x^2\right) = Q(x,y,z)\,. \end{aligned}$$

Lema 7.2 *Sean* $P(x,y,z)$ *y* $Q(x,y,z)$ *dos polinomios homogéneos equivalentes. Entonces sus curvas algebraicas proyectivas* $\tilde{C}(P)$ *y* $\tilde{C}(Q)$ *son equivalentes.*

Demostración. Por hipótesis tenemos $t \neq 0$ y una transformación lineal $f : \mathbb{R}^3 \to \mathbb{R}^3$ tales que

$$Q = t\ (P \circ f)\,.$$

Abusando de la notación, pensamos que f también es función de $\mathbb{P}^2$ en sí mismo con la fórmula $f([\mathbf{x}]) = [f(\mathbf{x})]$. Demostraremos que $f\left(\tilde{C}(Q)\right) = \tilde{C}(P)$.

Supongamos primero que $[\mathbf{x}] \in \tilde{C}(Q)$, entonces $Q(\mathbf{x}) = 0$. Para ver si $f([\mathbf{x}]) = [f(\mathbf{x})] \in \tilde{C}(P)$, tenemos que evaluar a P en $f(\mathbf{x})$:

$$P(f(\mathbf{x})) = (P \circ f)(\mathbf{x}) = t^{-1} Q(\mathbf{x}) = 0.$$

Por lo tanto $f([\mathbf{x}]) \in \tilde{C}(P)$ y concluimos que $f\left(\tilde{C}(Q)\right) \subset \tilde{C}(P)$.

Supongamos ahora que $[\mathbf{y}] \in \tilde{C}(P)$, *i.e.* que $P(\mathbf{y}) = 0$; para ver que $[\mathbf{y}] \in f\left(\tilde{C}(Q)\right)$ debemos encontrar $[\mathbf{x}] \in \tilde{C}(Q)$ tal que $f([\mathbf{x}]) = [\mathbf{y}]$. Para esto, como f es transformación, tiene una inversa y podemos declarar $\mathbf{x} := f^{-1}(\mathbf{y})$. Como $\mathbf{y} \neq \mathbf{0}$ pues es un punto en $\mathbb{P}^2$ y f^{-1} es una transformación, $\mathbf{x} \neq \mathbf{0}$ y $[\mathbf{x}] \in \mathbb{P}^2$ está bien definido. Nos falta ver que $[\mathbf{x}]$ está en $\tilde{C}(Q)$, y para esto hay que evaluar

$$Q(\mathbf{x}) = t\ (P \circ f)(f^{-1}(\mathbf{y})) = t\,P(\mathbf{y}) = 0.$$

Por lo tanto $\tilde{C}(P) \subset f\left(\tilde{C}(Q)\right)$ y queda demostrado el lema. □

Sería muy bonito que el lema tuviera inverso, que fuera un teorema "si y sólo si". Pero no es estrictamente cierto. Por ejemplo, si P fuera el cuadrado de un polinomio y Q el cubo, definen la misma curva pero no son equivalentes. También tenemos el problema de que los reales no son algebraicamente cerrados. Por ejemplo, en $\mathbb{P}^1$ el polinomio homogéneo $x^2 + y^2$ define el vacío (su única solución real es $(0,0)$, que proyectivamente no vale), así que el producto $\left(x^2 + y^2\right)(x + y)$ y el polinomio homogéneo lineal $x + y$ definen al mismo punto $[1 : -1]$ pero no son proyectivamente equivalentes ni potencias de un mismo polinomio. Lo que sí es cierto a grandes rasgos es que, sobre los complejos, si dos polinomios definen la misma "curva" entonces, salvo constantes, son potencias de algún otro. Éste es un teorema famoso conocido como el "**Nullstellensatz**" que habla sobre los ceros de polinomios. De nuevo, rebasa el ámbito de este libro. Pero no nos desanimemos, en ciertos casos (el cuadrático y el lineal) sí daremos la demostración pero hay que desarrollar más la teoría geométrica.

EJERCICIO 7.7 Demuestra que la equivalencia proyectiva de subconjuntos es una relación de equivalencia.

EJERCICIO 7.8 Demuestra que la equivalencia de polinomios es una relación de equivalencia.

EJERCICIO 7.9 Demuestra que la hipérbola $x^2 - y^2 - 1$ es proyectivamente equivalente al círculo $x^2 + y^2 - 1$.

EJERCICIO 7.10 Demuestra que hay una única clase de equivalencia de rectas.

7.2 Formas cuadráticas

Basta de generalidades. Concentrémonos en el caso de las curvas cuadráticas o cónicas. Una curva cuadrática en $\mathbb{R}^2$ está determinada por un polinomio de segundo grado que podemos expresar como

$$P(x, y) = ax^2 + 2bxy + cy^2 + 2dx + 2ey + f.$$

(Nótese que añadimos un 2 a los términos lineales, a diferencia del Capítulo 4, pues en un momento, al proyectivizar, se comportarán como términos mixtos con la nueva variable.) Su homogeneización es

$$\tilde{P}(x, y, z) = ax^2 + 2bxy + cy^2 + 2dxz + 2eyz + fz^2.$$

Si tomamos un vector variable $\mathbf{x} = (x, y, z)$, entonces podemos simplificar la notación como en el Capítulo 4. Este polinomio se escribe

$$\tilde{P}(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} = \mathbf{x} \cdot A\mathbf{x},$$

donde A es la matriz simétrica

$$A = \begin{pmatrix} a & b & d \\ b & c & e \\ d & e & f \end{pmatrix},$$

pues (y aquí aparece la importancia de haber añadido 2 en los términos mixtos),

$$\begin{aligned} \mathbf{x}^\top A\mathbf{x} &= (x, y, z) \begin{pmatrix} a & b & d \\ b & c & e \\ d & e & f \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= (x, y, z) \begin{pmatrix} ax + by + dz \\ bx + cy + ez \\ dx + ey + fz \end{pmatrix} \\ &= x(ax + by + dz) + y(bx + cy + ez) + z(dx + ey + fz) \\ &= \tilde{P}(x, y, z). \end{aligned}$$

Llamaremos *forma cuadrática* a un polinomio homogéneo de grado 2. Por lo que acabamos de ver, la forma cuadrática general se escribe como

$$\mathsf{P}(\mathbf{x}) = \mathbf{x}^\top \mathsf{A}\mathbf{x} \quad \text{con} \quad \mathsf{A} = \mathsf{A}^\top; \tag{7.7}$$

y la única posibilidad de que no sea la homogeneización de un polinomio cuadrático en dos variables es que tenga la forma $2dxz + 2eyz + z^2 = z(2dx + 2ey + z)$, cuya curva es un par de rectas y siendo una de ellas la del infinito. Así que conviene considerar el caso de una curva cuadrática proyectiva definida por una forma cuadrática en tres variables como en (7.7), donde le hemos quitado ya la tilde a P para simplificar nuestra notación.

La matriz simétrica A que define una forma cuadrática es el paquete que lleva toda su información: los coeficientes. Veamos ahora cómo se traduce la equivalencia de polinomios en las matrices simétricas que los determinan. Tenemos que ver qué le pasa cuando precomponemos con una transformación lineal $f : \mathbb{R}^3 \to \mathbb{R}^3$ que está definida como $f(\mathbf{x}) = \mathsf{B}\mathbf{x}$ donde B es una matriz no singular. Se tiene

$$\begin{aligned}(\mathsf{P} \circ f)(\mathbf{x}) &= \mathsf{P}(f(\mathbf{x})) = \mathsf{P}(\mathsf{B}\mathbf{x}) \\ &= (\mathsf{B}\mathbf{x})^\top \mathsf{A}(\mathsf{B}\mathbf{x}) = \left(\mathbf{x}^\top \mathsf{B}^\top\right) \mathsf{A}(\mathsf{B}\mathbf{x}) \\ &= \mathbf{x}^\top \left(\mathsf{B}^\top \mathsf{A}\mathsf{B}\right)\mathbf{x}.\end{aligned}$$

Lo cual implica que dos matrices simétricas A y A' determinan formas cuadráticas equivalentes si existen $t \neq 0$ y $\mathsf{B} \in \mathbf{GL}(3)$ tales que

$$\mathsf{A}' = t\left(\mathsf{B}^\top \mathsf{A}\mathsf{B}\right);$$

diremos en este caso que A y A' son *algebraicamente equivalentes*.

El problema de clasificar curvas cuadráticas proyectivas se reduce entonces a encontrar las formas canónicas a las que podemos reducir matrices algebraicamente equivalentes. Pero antes de entrar de lleno en él, concluyamos un ejemplo importante que dejamos pendiente: el de la parábola.

Sean P y su homogeneización $\tilde{\mathsf{P}}$ como en (7.5) y (7.6) respectivamente. Dicho de otra manera, su matriz asociada es

...Dibujo

$$\mathsf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}.$$

Sabemos que $\mathsf{C}(\tilde{\mathsf{P}})$ es un cono circular (ver Ejercicio 7.4) y además que es tangente al plano $z = 0$ (y también al $y = 0$). Isométricamente es igual al cono del círculo unitario; y la isometría que lleva a este último en $\mathsf{C}(\tilde{\mathsf{P}})$ es la rotación de $-45°$ alrededor del eje x. Sea B la matriz de esa rotación, es decir

$$\mathsf{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{pmatrix}.$$

Tenemos entonces, de acuerdo con la demostración del Lema 7.2, que

$$\begin{aligned} B^\top AB &= \frac{1}{2}\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 1 \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & -1 \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \end{aligned}$$

es la matriz del cono de $\mathbb{S}^1$, al que llamaremos el *cono de Lorentz.*

Junto con el Ejercicio 7.9, tenemos entonces que la hipérbola, la parábola y el círculo son proyectivamente equivalentes. Lo que afinmente las diferencía es cómo intersectan a la línea al infinito: la hipérbola en dos puntos (correspondientes a la dirección de sus asíntotas), la parábola en uno (correspondiente a su directriz) y el círculo en ninguno.

7.2.1 Clasificación usando a la afín

Puesto que el grupo afín $\mathbf{Af}(2)$ se puede considerar naturalmente dentro del grupo general lineal, $\mathbf{Af}(2) \subset \mathbf{GL}(3)$, como las matrices de la forma

$$\begin{pmatrix} B' & \mathbf{k} \\ 0^\top & 1 \end{pmatrix}, \tag{7.8}$$

donde $B' \in \mathbf{GL}(2)$ (la parte lineal) y $\mathbf{k} \in \mathbb{R}^2$ (la traslación), actuando en el plano $z = 1$ (ver Sección 6.6.4); entonces podemos usar la clasificación afín del Capítulo 4 para dar la clasificación proyectiva. Esta clasificación es tanto de formas cuadráticas, como de sus curvas algebraicas y sus matrices simétricas asociadas. Enunciémoslo en términos de matrices, aunque en la demostración hablaremos indistintamente de estos tres tipos de objetos hermanados.

Teorema 7.3 *Una matriz simétrica* 3×3 *distinta de cero es algebraicamente equivalente a una y sólo una de las siguientes*

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Demostración. Sabemos que afinmente se reducen a nueve (ver Sección 4.2), es decir, que actuando únicamente con matrices afines (7.8) llegamos a nueve formas canónicas. Pero ya hemos visto que tres de ellas (círculo, parábola e hipérbola) son proyectivamente equivalentes (para hacerlas equivalentes se usan necesariamente matrices no afines). De tal manera que las *no degeneradas* de la primera línea son dos: la *forma cuadrática definida* (o *euclidiana*) cuya curva es vacía, y la *forma cuadrática de Lorentz* que define a **la** cónica proyectiva no degenerada (representada por la forma cuadrática del círculo unitario).

Las formas *degeneradas* son las tres del segundo renglón. La primera tiene como curva un único punto, el $[0:0:1]$. En ella se han fundido lo que habíamos llamado "el círculo de radio cero" (x^2+y^2) y "las rectas paralelas imaginarias" (x^2+1), que, al homogeinizarse en x^2+z^2, sí tiene un punto al infinito. La segunda es un par de rectas (antes eran dos casos: paralelas y no que en el proyectivo son lo mismo). Y por último, una recta doble.

Puesto que los subconjuntos del proyectivo asociados a estas matrices (las curvas de sus respectivas formas) son todos de distinta índole, no es posible que sean equivalentes. Y esto concluye la demostración. □

Podría parecer que hemos acabado nuestro trabajo de clasificación de curvas cuadráticas proyectivas, y en principio así es. Pero repasemos la demostración que dimos en el fondo. La clasificación afín. Consistió en reducir el polinomio general a alguno de los canónicos en dos pasos centrales pero independientes. Primero, encontrar un centro (para eliminar términos lineales) por medio de una traslación, y, segundo, encontrar los ejes (para eliminar términos mixtos) por medio de una rotación. Lo demás fueron detalles, se reescalaron los ejes coordenados y se multiplicó por la constante si fuera requerido para hacer de todos los coeficientes 0, 1 o -1; y finalmente, se permutaron variables.

Al proyectivizar y homogeneizar el polinomio nos hemos "deshecho" de los términos lineales y el constante; ahora todo es cuadrático, entonces parece que podríamos evitar el primer paso, y hacerlo todo de golpe con la rotación. Esto nos dará una demostración más elegante, cuyas ventajas serán múltiples. Por un lado nos dará una idea general de cómo deben ser las superficies cuadráticas en todas las dimensiones, y por el otro, un entendimiento más profundo de las formas cuadráticas que redundará en conocimientos concretos sobre las cónicas y en nuevas geometrías.

7.2.2 Equivalencia lineal

Sea A una matriz simétrica 3×3 y P su forma cuadrática asociada, es decir,

$$P(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} = \mathbf{x}\cdot(A\mathbf{x}) \quad \text{con} \quad A^\top = A.$$

Para entender con más precisión qué hace esta forma cuadrática en $\mathbb{R}^3$, y en particular cómo es la geometría de su cono asociado $C(P)$, actuaremos en ella, no a lo bestia

con matrices cualesquiera de $\mathbf{GL}(3)$ como en la sección anterior, sino con más tacto, con las que preservan el producto interior y por tanto la métrica. Sabemos que estas matrices son las matrices ortogonales (que forman el grupo $\mathbf{O}(3)$) y que también preservan la métrica de $\mathbb{P}^2$.

Si $B \in \mathbf{O}(3)$, y $f : \mathbb{R}^3 \to \mathbb{R}^3$ es su transformación asociada, tenemos que

$$(P \circ f)(\mathbf{x}) = \mathbf{x}^\top \left(B^\top A B\right) \mathbf{x},$$

como para cualquier matriz B. Pero en este caso especial ($B \in \mathbf{O}(3)$), la matriz $B^\top AB$ tiene una interpretación geométrica muy particular. Puesto que

$$B \in \mathbf{O}(3) \quad \Leftrightarrow \quad B^\top = B^{-1},$$

entonces $B^\top AB = B^{-1}AB$ y esta matriz es la matriz asociada a la transformación A pero escrita en la base B. ...Diujo

Veámoslo. Si B tiene columnas $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, entonces el que sea invertible significa que estos tres vectores forman una base, es decir, que cualquier vector $\mathbf{x} \in \mathbb{R}^3$ se puede escribir como

$$\mathbf{x} = \beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \beta_3 \mathbf{u}_3. \tag{7.9}$$

Definimos $\mathbf{x}_B := (\beta_1, \beta_2, \beta_3)$ como la expresión de $\mathbf{x}$ *en la base* B. Así que, por ejemplo, $\mathbf{u}_1$ en la base B es $(1, 0, 0) = \mathbf{e}_1$; pero $\mathbf{u}_1$ en la base canónica puede ser complicado. Para pasar de la expresión en la base B a la expresión en la base canónica, simplemente hay que multiplicar por B, pues equivale a calcular los valores reales de las coordenadas en la expresión (7.9), que se puede reescribir como $\mathbf{x} = B\mathbf{x}_B$. Y al revés, para pasar de la expresión en la base canónica a la expresión en la base B hay que multiplicar por la inversa B^{-1}, pues es la manera de encontrar la solución de (7.9) pensado como sistema de ecuaciones con incógnitas $\beta_1, \beta_2, \beta_3$; puesto que $\mathbf{x} = B\mathbf{x}_B$, tenemos que $\mathbf{x}_B = B^{-1}\mathbf{x}$. En el ejemplo, está claro que $B^{-1}\mathbf{u}_1 = \mathbf{e}_1$. Así que la matriz $B^{-1}AB$ pensada como transformación aplicada a un vector escrito en la base B, primero lo reescribe en la base canónica, luego le aplica A (que es como ella sabe actuar) y luego lo traduce de nuevo a la base B. Con la notación que desarrollamos en el párrafo anterior, esto se escribe

$$\left(B^{-1}AB\right)\mathbf{x}_B = B^{-1}A\left(B\mathbf{x}_B\right) = B^{-1}A\mathbf{x} = B^{-1}\left(A\mathbf{x}\right) = \left(A\mathbf{x}\right)_B.$$

Hemos demostrado que $B^{-1}AB$ es la matriz de la transformación A escrita en la base B. Es natural entonces decir que dos matrices son *linealmente equivalentes* si expresan la misma transformación pero en diferentes bases, es decir si son A y $B^{-1}AB$ para alguna $B \in \mathbf{GL}(3)$, pues obsérvese que en nuestra discusión de la escritura en diferentes bases sólo usamos que B fuera invertible. Si, en particular, $B \in \mathbf{O}(3)$, que es el caso que nos interesa, entonces se dice, aún más, que A y $B^{-1}AB$ son *ortogonalmente equivalentes.* Dibujo...

Hay que remarcar que la equivalencia lineal **no** corresponde a la equivalencia algebraica que estamos estudiando (de ahí que haya sido necesario poner apellidos),

pues en general $B^\top AB \neq B^{-1}AB$. Sólo en el caso especial en que se actúe con las ortogonales (cuando $B \in \mathbf{O}(3)$, y que cumple que su inversa es su transpuesta), sí coinciden. Es decir...

Lema 7.4 *Si dos matrices simétricas son ortogonalmente equivalentes, entonces son algebraicamente equivalentes.* □

Podemos entonces pensar a A como función lineal y no como forma cuadrática para simplificarla ortogonalmente.

7.3 Diagonalización de matrices simétricas

Esta sección está dedicada a demostrar lo siguiente.

Teorema 7.5 *Sea* A *un matriz simétrica. Entonces es ortogonalmente equivalente a una matriz diagonal, es decir, existe una matriz ortogonal* B *tal que*

$$B^\top AB = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}. \tag{7.10}$$

Y para demostrarlo podemos pensar a A como función lineal (o proyectiva) pues como $B \in \mathbf{O}(3)$, $B^\top = B^{-1}$ y entonces la matriz $B^\top AB$ representa a la función A escrita en otra base. Que en esta nueva base sea diagonal, como vimos en la demostración del Teorema de 4 en 4, significa que, proyectivamente, fija los tres puntos de la base; y euclidianamente, que alarga (o acorta) a sus ejes, que los deja en sí mismos. Para encontrar a B buscamos entonces puntos fijos de la función A (proyectiva); o bien, vectores propios de la función lineal.

Recordemos de la Sección 4.3.2, como de hecho gran parte de la discusión anterior ya lo hace, que un *vector propio* de la matriz A es un $\mathbf{v} \in \mathbb{R}^3$ distinto de $\mathbf{0}$ para el cual existe una $\lambda \in \mathbb{R}$, su *valor propio*, tal que

$$A\mathbf{v} = \lambda\mathbf{v},$$

y a la pareja $\mathbf{v}, \lambda$ la llamaremos un *par propio* de A. Si $\lambda \neq 0$, $[\mathbf{v}]$ es un punto fijo de A (pensada como transformación proyectiva), y si $\lambda = 0$, es un "punto de proyección" donde no está definido su valor. También vimos en 4.3.2 que para encontrar pares propios, primero se encuentra un valor propio como raíz del polinomio característico y luego se encuentra una solución no trivial al sistema de ecuaciones lineales correspondiente. Como en aquel entonces trabajábamos en $\mathbb{R}^2$, reconstruyamos la argumentación en $\mathbb{R}^3$ haciendo énfasis en qué se necesitaría para el caso general.

Lema 7.6 *Dado* $\lambda \in \mathbb{R}$, *entonces existe* $\mathbf{v} \neq \mathbf{0}$ *tal que* $A\mathbf{v} = \lambda\mathbf{v}$ *si y sólo si* $\det(A - \lambda I) = 0$.

Demostración. Tenemos que

$$\begin{aligned} & A\boldsymbol{v} = \lambda \boldsymbol{v} \\ \Leftrightarrow\ & A\boldsymbol{v} - \lambda\boldsymbol{v} = 0 \\ \Leftrightarrow\ & (A - \lambda I)\,\boldsymbol{v} = 0. \end{aligned}$$

Y la última ecuación sucede con algún $\boldsymbol{v} \neq 0$ si y sólo si $\det(A - \lambda I) = 0$ (ver el Teorema 5.9 para el caso en $\mathbb{R}^3$, y esto es lo que se necesita demostrar en el caso general). □

Dibujos...

Regresemos al ejemplo de la parábola con esta óptica. Si A es la matriz de la parábola, $\det(A - \lambda I)$ toma la forma

$$\begin{aligned} \det\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}\right) &= \det\left(\begin{pmatrix} 1-\lambda & 0 & 0 \\ 0 & -\lambda & -1 \\ 0 & -1 & -\lambda \end{pmatrix}\right) \\ &= (1-\lambda)\left(\lambda^2 - 1\right) \\ &= -(\lambda - 1)^2(\lambda + 1). \end{aligned}$$

Lo cual nos dice que los valores propios de A son 1 y -1 pues son las *raíces* de este polinomio (los valores de λ que lo anulan). Para el valor propio -1, se tiene entonces que el sistema de ecuaciones homogéneas

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

tiene soluciones no triviales, a saber cualquiera de la forma $(0, t, t)$ con $t \neq 0$, correspondientes al vector propio $\mathbf{u} := (0, 1, 1)$ ($A\mathbf{u} = -\mathbf{u}$) y al eje de simetría del cono. Por su parte, el valor propio 1, que tiene multiplicidad 2, da el sistema

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & -1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

cuyas soluciones son todos los puntos del plano $y + z = 0$; que es el plano de simetría del cono. Proyectivamente, la recta $y + z = 0$, que es precisamente $\ell_{[\mathbf{u}]}$ se queda fija punto a punto, de tal manera que la matriz A representa la reflexión en $\ell_{[\mathbf{u}]}$, que también es la inversión en su polar $[\mathbf{u}]$. ...Dibujo

Regresando al caso general, el lema nos dice que los valores propios de la matriz A son las raíces (los ceros) de su *polinomio característico*

$$\det(A - \lambda I),$$

donde se piensa a λ como una variable. El polinomio característico de una matriz no cambia cuando la escribimos en otra base, pues para cualquier $B \in \mathbf{GL}(3)$:

$$\begin{aligned}
\det\left(B^{-1}AB - \lambda I\right) &= \det\left(\left(B^{-1}AB\right) - \lambda\left(B^{-1}B\right)\right) \\
&= \det\left(\left(B^{-1}AB\right) - \left(B^{-1}(\lambda I)\,B\right)\right) \\
&= \det\left(B^{-1}(A - \lambda I)\,B\right) \\
&= \det\left(B^{-1}\right)\det(A - \lambda I)\det(B) \\
&= \det(A - \lambda I)\,.
\end{aligned}$$

Así que diagonalizar una matriz (como en 7.10) incluye necesariamente encontrar las raíces de su polinomio característico, pues entonces se tiene por lo anterior que

$$\begin{aligned}
\det(A - \lambda I) &= (\lambda_1 - \lambda)(\lambda_2 - \lambda)(\lambda_3 - \lambda) \\
&= -(\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3)\,,
\end{aligned}$$

que es escribir el polinomio en términos de sus raíces.

Si A es una matriz 3×3 (como en el caso que nos interesa), éste es un polinomio de grado 3, y por el Teorema Fundamental del Algebra,[2] tiene al menos una raíz real. En realidad lo que dice este teorema es que tiene tres raices complejas, pero si sus coeficientes son reales, como en nuestro caso, entonces las raíces complejas vienen en parejas; un complejo y su conjugado. Como el grado es impar, entonces hay al menos una raíz real. Hemos demostrado entonces lo siguiente.

Corolario 7.7 *Toda matriz de* 3×3 *tiene al menos un par propio.* □

Ahora, usaremos la simetría de A para aplicar inducción encontrando un plano "invariante" y buscar ahí los dos pares propios que nos faltan. En general, dada una matriz cuadrada A se dice que un subespacio V es *invariante* si se cumple que

$$\mathbf{x} \in V \quad \Rightarrow \quad A\mathbf{x} \in V.$$

Es decir si la función A lo deja en su lugar (aunque posiblemente cambiado internamente). El ejemplo básico de espacios invariantes corresponden a las líneas generadas por los vectores propios, pues la condición $A\mathbf{v} = \lambda\mathbf{v}$ implica que $[\mathbf{v}]$ es invariante y viceversa, una línea (por el origen) invariante consta de vectores propios.

Recordemos del Capítulo 4 la caracterización de las matrices simétricas como aquellas que "saltan" en el producto interior. Por su sencillez, y para que quede autocontenida nuestra discusión, lo volveremos a demostrar.

Lema 7.8 *Sea* A *una matriz de* $n \times n$. *Entonces*

$$A = A^\top \quad \Leftrightarrow \quad (A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y}) \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

[2] Ver Apéndice.

Demostración. Para la ida ($\Rightarrow$) se usa que el producto interior es un caso particular del producto de matrices y su asociatividad. Para cualquier $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$(A\mathbf{x}) \cdot \mathbf{y} = (A\mathbf{x})^\top \mathbf{y} = \left(\mathbf{x}^\top A^\top\right) \mathbf{y} = \left(\mathbf{x}^\top A\right) \mathbf{y} = \mathbf{x}^\top (A\mathbf{y}) = \mathbf{x} \cdot (A\mathbf{y}).$$

Para demostrar el otro lado ($\Leftarrow$), basta ver que la entrada i, j de A es $\mathbf{e}_i \cdot (A\mathbf{e}_j)$. Por hipótesis, $\mathbf{e}_i \cdot (A\mathbf{e}_j) = \mathbf{e}_j \cdot (A\mathbf{e}_i)$, que es la entrada j, i, así que A es simétrica. □

Hay que hacer resaltar que la condición de ser simétrica, para una transformación, se ha expresado en términos del producto interior. Así que si tiene un subespacio invariante, se puede decir cuándo es simétrica en él y éste será el fundamento de la inducción después de encontrar el plano invariante.

Lema 7.9 *Sea A una matriz simétrica ($A = A^\top$) y $\mathbf{v}$ un vector propio de A. Entonces el plano ortogonal a $\mathbf{v}$ es un espacio invariante de A; es decir*

$$\mathbf{x} \cdot \mathbf{v} = 0 \quad \Rightarrow \quad (A\mathbf{x}) \cdot \mathbf{v} = 0.$$

Demostración. Si $\mathbf{v}$ un vector propio de A se tiene que $A\mathbf{v} = \lambda\mathbf{v}$ para alguna $\lambda \in \mathbb{R}$, y entonces $\mathbf{x} \cdot \mathbf{v} = 0$ implica que

$$(A\mathbf{x}) \cdot \mathbf{v} = \mathbf{x} \cdot (A\mathbf{v}) = \mathbf{x} \cdot (\lambda\mathbf{v}) = \lambda (\mathbf{x} \cdot \mathbf{v}) = 0.$$

□

Ya tenemos todos los ingredientes para demostrar el Teorema de diagonalización ortogonal de matrices simétricas.

Sea A una matriz simétrica 3×3. Por el corolario 7.7, A tiene un valor propio λ_3. Sea $\mathbf{u}_3$ un vector propio unitario correspondiente a λ_3 (cualquier vector propio multiplicado por el inverso de su magnitud lo hace unitario). Sean $\mathbf{v}_1$ y $\mathbf{v}_2$ cualquier par de vectores tales que $\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_3$ forman una base ortonormal. Puesto que $\mathbf{v}_1$ y $\mathbf{v}_2$ son ortogonales a $\mathbf{u}_3$, el Lema 7.9 nos dice que $A\mathbf{v}_1$ y $A\mathbf{v}_2$ también lo son, de tal manera que A escrita en la base $\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_3$ tiene la forma

$$A' = \begin{pmatrix} a & b & 0 \\ b & c & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix},$$

pues $b = \mathbf{v}_1 \cdot A\mathbf{v}_2 = \mathbf{v}_2 \cdot A\mathbf{v}_1$. Como lo demostramos en el Capítulo 4, se puede ver directamente que el polinomio característico de esta matriz tiene sus otras dos raíces reales (ya nada más falta resolver un polinomio cuadrático). Que, como no cambia el polinomio característico, son valores propios de la matriz original. Sean λ_1 y λ_2 estas raíces y sean $\mathbf{u}_1$ y $\mathbf{u}_2$ sus vectores propios unitarios correspondientes. Nos falta ver que son ortogonales.

Por ser A simétrica,

$$\lambda_i (\mathbf{u}_i \cdot \mathbf{u}_j) = (\lambda_i \mathbf{u}_i) \cdot \mathbf{u}_j = (A\mathbf{u}_i) \cdot \mathbf{u}_j = \mathbf{u}_i \cdot (A\mathbf{u}_j) = \mathbf{u}_i \cdot (\lambda_j \mathbf{u}_j) = \lambda_j (\mathbf{u}_i \cdot \mathbf{u}_j),$$

lo que implica que $(\lambda_i - \lambda_j)(\mathbf{u}_i \cdot \mathbf{u}_j) = 0$. Por tanto, tenemos que *si* $\lambda_i \neq \lambda_j$ *entonces* $\mathbf{u}_i$ *es ortogonal a* $\mathbf{u}_j$. Si dos valores propios coincidieran, podemos suponer que son λ_1 y λ_2, pero entonces la matriz A' ya sería diagonal pues A en el plano invariante es una homotecia. Y si los tres valores propios coinciden, entonces A es una homotecia en $\mathbb{R}^3$ y desde el principio es diagonal.

Con esto se concluye la demostración del Teorema de diagonalización 7.5.

7.4 Geometría de las formas cuadráticas

Resumiendo las dos secciones anteriores, a una forma cuadrática en tres variables le asociamos una matriz simétrica A y hemos demostrado que en una base ortonormal adecuada $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, ésta se escribe como una matriz diagonal

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

Lo cual dice que, en la nueva base, la forma cuadrática se expresa de manera muy simple:

$$P(\mathbf{x}) = \lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2,$$

Dibujo... donde (x, y, z) son ahora las coordenadas de $\mathbf{x}$ respecto a la base $B = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$. Se tiene además que el determinante de la matriz original es el producto de sus valores propios ($\det(A) = \lambda_1\lambda_2\lambda_3$). Entonces se puede saber desde el principio si alguno de los valores propios es cero, simplemente corroborando si $\det(A)$ es cero. A este caso se le llama *degenerado* pues, en la base adecuada, no depende de alguna de las coordenadas; viene, por decirlo así, de algo en una dimensión más baja y hay al menos una variable de pacotilla; es un "cilindro" sobre los ceros de una forma cuadrática de dimensión menor.

El caso interesante es el *no degenerado*, cuando $\det(A) \neq 0$. Y éste se divide en dos subcasos. Primero, cuando los tres valores propios tienen el mismo signo es llamada forma cuadrática *definida*, pues el valor de la forma cuadrática siempre tendrá el mismo signo (excepto en el $\mathbf{0}$ que vale 0). Si fueran positivos, por ejemplo, no importa qué valor le demos a las variables, al elevarlas al cuadrado se harán positivas y estaremos sumando tres números positivos; se le llama entonces *positiva definida*. Cuando los tres son negativos se le llama *negativa definida*. Puesto que el único cero de las formas cuadráticas definidas es el origen, su curva proyectiva asociada es vacía. ...Dibujo

La otra posibilidad para una forma cuadrática no degenerada es que entre los valores propios aparezcan los dos signos, y se le llamará *indefinida*. Dos de los valores

propios son de un signo y el restante del contrario; podemos suponer (cambiando el orden de la base si fuera necesario) que los dos primeros λ_1 y λ_2 son del mismo signo, y además, para fijar ideas, que son positivos (el otro caso será equivalente si se intercambian las palabras "positivo" y "negativo"). Éstas son las formas cuadráticas que definen a las *cónicas proyectivas*. En $\mathbb{R}^3$ sus ceros son un cono sobre una elipse. Demostrémoslo. Su intersección con el plano $z = 1$ (recordemos que estamos en la base B y no necesariamente en la canónica) está determinada por la ecuación

$$\lambda_1 x^2 + \lambda_2 y^2 = (-\lambda_3)\,,$$

donde las tres constantes son positivas, y por lo tanto es una elipse. Entonces $C(P)$ consiste de todas las rectas por el origen y puntos de esta elipse. Dibujo...

7.4.1 Su simetría

Los tres ejes $[\mathbf{u}_1], [\mathbf{u}_2], [\mathbf{u}_3]$ de la forma cuadrática rigen su simetría (y al revés). Como hemos visto, se puede definir un subgrupo de $\mathbf{O}(3)$ para cualquier figura que son sus simetrías. También se puede definir para funciones en general, pero en nuestro caso particular definámoslo para la forma cuadrática que estamos estudiando:

$$\mathbf{Sim}(P) = \{f \in \mathbf{O}(3) \mid P = P \circ f\}.$$

Aquí consideramos a los elementos de $\mathbf{O}(3)$ como funciones, y estamos tomando aquellas que, al precomponerlas con la forma cuadrática, ésta ni se entera. Cualquier cosa que definamos con base en P, por ejemplo sus ceros o sus superficies de nivel que estudiaremos más adelante, heredarán esta simetría.

Si llamamos f_i a la reflexión en el plano generado por $\mathbf{u}_j$ y $\mathbf{u}_k$ donde $\{i, j, k\} = \{1, 2, 3\}$, entonces se escribe en coordenadas (respecto a B, insistimos) cambiando el signo a la coordenada i. Por ejemplo $f_1(x, y, z) = (-x, y, z)$. Está claro entonces que $f_i \in \mathbf{Sim}(P)$ pues como la coordenada en cuestión se eleva al cuadrado al evaluar P, el cambio de signo no altera el resultado. Cuando los tres valores propios λ_1, λ_2 y λ_3 son diferentes, entonces f_1, f_2 y f_3 generan a $\mathbf{Sim}(P)$ y este tiene ocho elementos. Las rectas $[\mathbf{u}_1], [\mathbf{u}_2]$ y $[\mathbf{u}_3]$ están entonces determinadas por la geometría —por la simetría— y por ello se les llama *ejes de simetría*.

Cuando dos de los valores propios coinciden, digamos que $\lambda_1 = \lambda_2$, entonces el grupo de simetrías crece. El plano generado por $\mathbf{u}_1$ y $\mathbf{u}_2$ es un espacio propio; cualquier vector ahí es vector propio para el valor propio λ_1. Entonces $\mathbf{u}_1$ y $\mathbf{u}_2$ pueden escogerse como cualquier base ortonormal ahí y habrá un subgrupo $\mathbf{O}(2)$ de $\mathbf{Sim}(P)$. En el caso de una forma cuadrática no degenerada indefinida ésta es la simetría máxima a la que puede aspirar (como en la forma cuadrática de Lorentz). De cualquier manera el *plano de simetría* queda determinado.

Cuando los tres valores propios coinciden (y por lo tanto es definida la forma cuadrática, de hecho una constante por la forma cuadrática euclidiana) la simetría es todo $\mathbf{O}(3)$.

EJERCICIO 7.11 Sean A una matriz simétrica 3×3 y $B \in \mathbf{O}(3)$ una matriz ortogonal. Si P es la forma cuadrática asociada a A y f la transformación lineal asociada a B ($f(\mathbf{x}) = B\mathbf{x}$), demuestra que $P = P \circ f$ si y sólo si $A = B^{-1}AB$.

EJERCICIO 7.12 Sean $A = (\lambda_1 \boldsymbol{e}_1, \lambda_2 \boldsymbol{e}_2, \lambda_3 \boldsymbol{e}_3)$ una matriz diagonal y $B = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) \in \mathbf{O}(3)$ una matriz ortogonal, tales que $A = B^\top AB$. Demuestra que si λ_1, λ_2 y λ_3 son diferentes, entonces $\mathbf{v}_i = \pm \boldsymbol{e}_i$ para $i = 1, 2, 3$.

Ejemplo

Consideremos un círculo unitario desplazado una unidad en el eje y; estudiaremos su cono por el origen al pensarlo en la carta coordenada canónica. El círculo está dado por el polinomio

$$P(x, y) = x^2 + (y - 1)^2 - 1 = x^2 + y^2 - 2y.$$

Dibujo... Al homogeneizar, su matriz asociada es

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 0 \end{pmatrix},$$

cuyo polinomio característico resulta ser

$$\det(A - \lambda I) = (1 - \lambda)^2 (-\lambda) - (1 - \lambda) = -\lambda^3 + 2\lambda^2 - 1.$$

En general no supondremos que sabemos resolver polinomios de tercer grado, pero en este caso, de la expresión intermedia está claro que se puede factorizar el factor lineal $(1 - \lambda)$ para obtener

$$\det(A - \lambda I) = (1 - \lambda)\left(\lambda^2 - \lambda - 1\right).$$

Tenemos entonces que una raíz del polinomio característico (o valor propio de A) es 1. Su vector propio asociado es $\boldsymbol{e}_1 = (1, 0, 0)$, que corresponde al hecho de que al trasladar el círculo en el eje y, el plano de simetría del cono gira sobre el eje x. Las otras dos raíces son:

$$\frac{1 \pm \sqrt{5}}{2}.$$

El valor propio $\phi = \left(1 + \sqrt{5}\right)/2$ es la *razón áurea* y el que falta es la no tan famosa solución negativa del polinomio áureo $\lambda^2 - \lambda - 1$, que resulta ser $-\phi^{-1}$ pues

$$\left(\frac{1 + \sqrt{5}}{2}\right)\left(\frac{1 - \sqrt{5}}{2}\right) = \frac{1 - 5}{4} = -1.$$

...Dibujos Para encontrar un vector propio para ϕ debemos resolver el sistema

$$\begin{pmatrix} 1-\phi & 0 & 0 \\ 0 & 1-\phi & -1 \\ 0 & -1 & -\phi \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

que nos da, usando el truco del compadre ortogonal en el segundo renglón de la matriz, la solución

$$\mathbf{u} := (0, 1, 1-\phi)\,.$$

Efectivamente, del polinomio áureo obtenemos que $\phi^{-1} = \phi - 1$, y por lo tanto

$$A\mathbf{u} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1-\phi \end{pmatrix} = \begin{pmatrix} 0 \\ \phi \\ -1 \end{pmatrix} = \phi \begin{pmatrix} 0 \\ 1 \\ -\phi^{-1} \end{pmatrix} = \phi\,\mathbf{u}.$$

Así que el plano de simetría del cono tiene pendiente $-\phi^{-1}$. El tercer vector propio, asociado al valor propio $-\phi^{-1}$, se puede obtener de manera análoga, o bien por el producto cruz (pues sabemos que son ortogonales); es $\boldsymbol{v} := \boldsymbol{e}_1 \times \mathbf{u} = \left(0, \phi^{-1}, 1\right)$. Obsérvese que el eje de simetría del cono no pasa por el centro del círculo original que es $(0, 1, 1)$. Este ejemplo demuestra que el centro, aunque es un invariante afín de las cónicas, no es un invariante proyectivo, es decir, proyectivamente no está bien definido. Con una simple traslación, el eje de simetría del cono no se traslado lo mismo, se "atraso" en proporción áurea. Y lo que era un cono circular se vuelve un cono elíptico. Dibujo...

Dibujo...

Por la teoría que desarrollamos, sabemos además que al cambiar coordenadas girando en el eje x para que el eje y caiga en $[\mathbf{u}]$ (y, por tanto, el eje z cae en $[\boldsymbol{v}]$), la forma cuadrática se expresa

$$x^2 + \phi y^2 - \phi^{-1} z^2.$$

Así que el origen "ve" al círculo (en el plano a distancia 1 de él y encuadrando la imagen) como la elípse

$$\phi\, x^2 + \phi^2\, y^2 = 1.$$

EJERCICIO 7.13 Considera el círculo unitario trasladado una distancia $t \geq 0$ a lo largo del eje y (en el ejemplo desarrollamos el caso $t = 1$). Encuentra los ejes de simetría del cono correspondiente y sus valores propios. Supón que el parámetro t representa el tiempo, ¿cómo se parametriza la familia de elipses que ve el origen si gira para seguir el movimiento del círculo? ¿Puedes demostrar que el único momento en que el centro del círculo se ve en el centro de la "elipse visual" es $t = 0$?

7.4.2 Reducción final

Teniendo el Teorema de Diagonalización de Matrices Simétricas en nuestras manos, podemos dar una demostración alternativa del Teorema de Clasificación 7.3. Simplemente se reduce a convertir una matriz diagonal

$$A = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

en una de ceros, unos y menos unos. Esto ya no se puede hacer con transformaciones ortogonales, pero se logra fácilmente reescalando los ejes coordenados. Suponiendo que A es no degenerada (las tres λ distintas de cero), la matriz de cambio de escalas que lo logra es

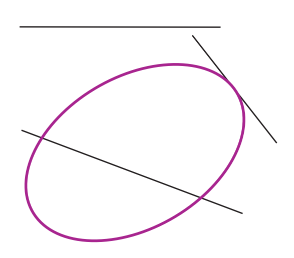
$$B = \begin{pmatrix} |\lambda_1|^{-\frac{1}{2}} & 0 & 0 \\ 0 & |\lambda_2|^{-\frac{1}{2}} & 0 \\ 0 & 0 & |\lambda_3|^{-\frac{1}{2}} \end{pmatrix}.$$

Pues es fácil ver que entonces $B^\top AB$ es, salvo permutación de las variables y signo, una de las dos matrices no degeneradas del Teorema 7.3.

7.5 Clasificación en $\mathbb{P}^3$ y en $\mathbb{R}^3$

7.5.1 Resumen de cónicas en $\mathbb{P}^2$ y en $\mathbb{R}^2$

En resumen, hemos demostrado que en el plano proyectivo hay una única cónica no degenerada y no vacía. Las tres cónicas clásicas en el plano afín aparecen porque hay tres tipos de líneas respecto a una cónica proyectiva dada: las que no la intersectan; las que la tocan en un sólo punto, llamadas sus *tangentes*, y las que la intersectan en dos puntos. Ello corresponde al hecho de que los polinomios cuadráticos en una sola variable tienen dos raíces complejas (conjugadas), una real de multiplicidad dos o bien dos raíces reales. Cuando una recta se manda al infinito para obtener una carta afín, dependiendo de qué tipo de recta sea respecto a una cónica dada, ésta se convierte respectivamente en elipse, parábola o hipérbola.



7.5.2 Dimensión 3

Sin entrar en detalles, describiremos la clasificación de superficies cuadráticas en $\mathbb{P}^3$ y en $\mathbb{R}^3$ que, a su vez, delinea una generalización natural a todas las dimensiones.

En $\mathbb{R}^3$, una superficie cuadrática es el conjunto de ceros de un polinomio de grado dos en tres variables. A éste se le hace corresponder un polinomio homogéneo de grado dos en cuatro variables cuyos ceros definen una superficie cuadrática en $\mathbb{P}^3$ que

es su cerradura proyectiva; agrega sólo puntos en el plano al infinito. A cualquier polinomio homogéneo de grado dos (o forma cuadrática) en cuatro variables se le asigna de manera única una matriz simétrica A de 4×4, de tal manera que se escribe

$$P(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x},$$

donde ahora $\mathbf{x}^\top = (x, y, z, w)$.

Clasificar proyectivamente las superficies cuadráticas equivale a encontrar formas canónicas a las cuales se puede reducir la matriz A utilizando la acción de $\mathbf{GL}(4)$ dada por $A \mapsto B^\top AB$ con $B \in \mathbf{GL}(4)$. Primero se estudia el caso particular de actuar por matrices ortogonales (en $\mathbf{O}(4)$), que cumplen $B^\top = B^{-1}$, y entonces esta acción equivale a escribir A en otra base ortonormal. Diagonalizar A de nuevo es el problema de encontrar valores y vectores propios. Análogamente a como procedimos, este problema se reduce a encontrar un valor propio real de A, o bien, algún subespacio invariante para aplicar lo que ya sabemos ahí (usar inducción). En esta dimensión se complica pues el polinomio característico es de grado par y no da en forma tan automática un valor propio. Pero si suponemos que las matrices simétricas son diagonalizables por cambio de base ortonormal (lo que además es cierto en cualquier dimensión aunque en este libro no nos toca demostrarlo), entonces prosiguiendo con un simple cambio de escala, una posible permutación de variables y, si fuera necesario, multiplicando por -1, obtenemos que la matriz A es equivalente a alguna de las siguientes:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

o bien es degenerada; es decir, una de las del Teorema 7.3 con una columna y un renglón más de puros ceros.

La primera corresponde a una forma cuadrática *definida* (sólo se anula en el origen de $\mathbb{R}^4$) que define al vacío en $\mathbb{P}^3$. La segunda define a la esfera unitaria en el $\mathbb{R}^3$ canónico, $w = 1$. Este es el caso análogo de la cónica no degenerada en $\mathbb{P}^2$, que podríamos llamar *elipsoide*. Pero tenemos un caso nuevo, la tercera que define un *hiperboloide reglado* y que abajo estudiaremos con más detenimiento. Así que

en $\mathbb{P}^3$ hay dos tipos de superficies cuadráticas no degeneradas no vacías.

La clasificación en $\mathbb{R}^3$ surgirá entonces de cómo se relaciona un plano, que mandaremos al infinito, con la superficie. Dibujos

Del elipsoide proyectivo surgen tres superficies en $\mathbb{R}^3$. Cuando el plano no la toca es un *elipsoide* común y corriente. Cuando el plano es tangente, se obtiene un *paraboloide* que se aproxima al infinito en una dirección dada, la *directriz*, y las secciones con planos ortogonales son elipses similares; en general, sus secciones por

planos son elipses excepto con los que contienen a la directriz que dan parábolas. Y finalmente, cuando el plano corta el elipsoide obtenemos el *hiperboloide de dos hojas,* cuyas secciones son hipérbolas o elipses.

Del hiperboloide reglado se obtienen dos superficies. Cuando el plano al infinito lo corta en una cónica no degenerada da el "equipal", conocido técnicamente como *hiperboloide de una hoja*. Y cuando el plano al infinito es tangente a él, lo intersecta en dos rectas (una cónica degenerada) y, entonces, vemos en el espacio afín un *paraboloide hiperbólico*.

7.5.3 Superficies regladas

Considerando los elipsoides como la generalización natural de las elipses, en dimensión 3 aparecen nuevas superficies cuadráticas que son las "*regladas*". Se llaman así porque tienen dos *reglados* que son conjuntos de rectas, o *reglas*, parametrizadas por una línea proyectiva. Consideremos primero el polinomio

Dibujo...

$$P(x, y, z) = x^2 + y^2 - z^2 - 1$$

que, en $\mathbb{R}^3$, define al *hiperboloide canónico* (de una hoja) que denotaremos $\mathcal{H}$.

Observemos primero que $\mathcal{H}$ es invariante bajo rotaciones en el eje z (es decir, que $\mathbf{O}(2) \subset \mathbf{Sim}(\mathcal{H})$). Pues el propio polinomio P es invariante bajo ellas. Se puede ver sustituyendo directamente las fórmulas en las variables del polinomio, o bien matricialmente (viendo que $B^\top AB = A$):

$$\begin{pmatrix} \cos\theta & \operatorname{sen}\theta & 0 & 0 \\ -\operatorname{sen}\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\theta & -\operatorname{sen}\theta & 0 & 0 \\ \operatorname{sen}\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

donde la matriz de enmedio, llamémosle A, es la matriz asociada a la homogeneización de P y las otras, $B^\top$ y B son las que representan las rotaciones en el eje z.

Observemos también que $\mathcal{H}$ contiene a las rectas

$$\ell_1 = \{\mathbf{e}_1 + t(\mathbf{e}_2 + \mathbf{e}_3) \mid t \in \mathbb{R}\} \quad \text{y} \quad \ell_2 = \{\mathbf{e}_1 + s(\mathbf{e}_3 - \mathbf{e}_2) \mid s \in \mathbb{R}\}\ , \quad \text{pues}$$

$$P(1, t, t) = 1 + t^2 - t^2 - 1 = 0 \quad \text{y} \quad P(1, -s, s) = 1 + s^2 - s^2 - 1 = 0.$$

Estas rectas, ℓ_1 y ℓ_2 , son las *reglas* de $\mathcal{H}$ que pasan por $\mathbf{e}_1$.

De estas dos observaciones se sigue que todas las rotaciones alrededor del eje z de la línea ℓ_1 (y de ℓ_2) están en $\mathcal{H}$; de hecho, "barren" a $\mathcal{H}$: al rotar ℓ_1 (o ℓ_2) se "genera" o se "pasa por" todo $\mathcal{H}$. Las rotaciones de ℓ_1 forman uno de los *reglados*, $\mathcal{R}_1$ denotémoslo; y las rotaciones de ℓ_2 forman el otro reglado $\mathcal{R}_2$. Se tiene que, proyectivamente, las reglas de uno y otro reglado siempre se intersectan, mientras que las reglas de un mismo reglado son ajenas entre sí (no se intersectan). De hecho, por cada punto de $\mathcal{H}$ pasan justo una regla de $\mathcal{R}_1$ y una de $\mathcal{R}_2$, que juntas generan el plano tangente a $\mathcal{H}$ en ese punto.

Para demostrar, más que simplemente enunciar, nuestras últimas aseveraciones conviene ver el hiperboloide $\mathcal{H}$ en su otra manifestación afín. Para esto, consideremos a $\mathcal{H}$ como el hiperboloide reglado proyectivo (es decir, en $\mathbb{P}^3$), definido por el polinomio homogéneo

$$\tilde{P}\,(x,y,z,w) = x^2 + y^2 - z^2 - w^2,$$

cuya matriz asociada A es la que ya usamos. Un cambio proyectivo de coordenadas, que geométricamente consiste, en $\mathbb{R}^4$, en rotar $45°$ en los planos xz y yw al mismo tiempo (y en $\mathbb{P}^3$ de medios giros en las rectas correspondientes), nos cambia a la matriz en

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\left(\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}\right) =$$

$$= \frac{1}{2}\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$



que corresponde al polinomio homogéneo $2\tilde{Q}$, donde

$$\tilde{Q}\,(x,y,z,w) = xz + yw.$$

Este polinomio es la homogeneización de

$$Q\,(x,y,z) = xz + y,$$

que define un *paraboloide hiperbólico,* o "silla de montar" pues parece tal al intercambiar las variables y con z. Sus dos reglados están claramente determinados por sus intersecciones con los planos $z = \mathrm{cte}$ y $x = \mathrm{cte}$; y que tienden al par de rectas $xz = 0$ en el plano al infinito.

***El hiperboloide por tres rectas**

...Dibujo

Una construcción elegante de superficies cuadráticas regladas con la que Hilbert y Chon-Vossen abren su libro clásico *Geometría e Imaginación* [9], es la siguiente.

Consideremos tres rectas ℓ_1, ℓ_2, ℓ_3 en $\mathbb{P}^3$ en *posición general*, *i.e.*, que no se intersecten dos a dos (aquí formalmente es importante estar pensando en términos proyectivos para excluir la posibilidad euclidiana del paralelismo y definir de un solo golpe la construcción). Para cada punto $p \in \ell_1$, tomamos el plano que genera junto con ℓ_2 y lo intersectamos con ℓ_3 para obtener un punto $q \in \ell_3$; la línea, ℓ_p, que pasa por p y q también intersecta a ℓ_2 pues son coplanares, así que intersecta a las tres líneas. Al variar $p \in \ell_1$ se obtiene un reglado de un hiperboloide $\mathcal{H}$, y ℓ_1, ℓ_2, ℓ_3 pertenecen a su otro reglado. Nótese que cualquier recta que intersecte a las tres dadas aparece como la ℓ_p de algún punto en ℓ_1, así que los hiperboloides reglados constan de todas las rectas que tocan a tres rectas dadas.

Es claro que en la construcción anterior pudimos haber utilizado las tres rectas en cualquier orden y obtener el mismo $\mathcal{H}$ con su reglado de "transversales" a ellas. Este reglado implica que, al tomar tres rectas en posición general en $\mathbb{P}^3$, se obtiene naturalmente una correspondencia entre ellas ($p_i \in \ell_i$ corresponde a $p_j \in \ell_j$ si están en una de las reglas de $\mathcal{H}$). Resulta que estas correspondencias entre líneas proyectivas son proyectivas. Así que otra manera de obtener hiperboloides reglados es con dos rectas ℓ_1 y ℓ_2 en posición general y una correspondencia proyectiva $f : \ell_1 \to \ell_2$, tomando todas las reglas por $p \in \ell_1$ y $f(p) \in \ell_2$. En el Capítulo 9 veremos algo similar a esto en el plano.

***EJERCICIO 7.14** Demuestra, o indica un esbozo de demostración de, que la unión de las líneas ℓ_p con $p \in \ell_1$ (que llamamos $\mathcal{H}$ arriba) es efectivamente una superficie cuadrática; es decir que está definido como los ceros de un polinomio homogéneo de grado 2.

***EJERCICIO 7.15** Delinea una demostración de que es proyectiva la correspondencia dada por el reglado entre las líneas ℓ_1, ℓ_2, ℓ_3 .

EJERCICIO 7.16 Suponiendo los dos ejercicios anteriores, demuestra que la construcción con dos líneas y una correspondencia proyectiva también es una superficie cuadrática.

7.5.4 *Idea de la clasificación general

Los métodos que hemos desarrollado se generalizan a todas las dimensiones y plantean naturalmente un programa para clasificar las "cuádricas". Señalaremos brevemente los pasos principales de este programa.

Se pueden considerar polinomios con n variables, $x_1, x_2, \cdots, x_n$, y a sus conjuntos de ceros se les llama *hipersuperficies algebraicas* en $\mathbb{R}^n$. Dentro de éstas, las definidas por polinomios de grado dos son las hipersuperficies cuadráticas, o simplemente *cuádricas*. La pregunta general es cúales son las posibles cuádricas en $\mathbb{R}^n$, y, de nuevo, hay que especificar un grupo de transformaciones para definir la equivalencia. El grupo natural a considerar es la generalización natural del afín a la dimensión correspondiente.

Añadiendo una nueva variable x_0, es fácil definir la homogeneización de tales

polinomios, y como sus ceros en $\mathbb{R}^{n+1}$ son unión de líneas por el origen estos definen las *hipersuperficies algebraicas proyectivas*, y su clasificación módulo el correspondiente grupo proyectivo es más sencilla y natural (sobre los complejos y otros campos, son objeto de estudio de la geometría algebraica). En particular, las *cuádricas proyectivas* están definidas por polinomios homogéneos de grado dos, o *formas cuadráticas*, y para ellos se tiene la misma expresión matricial, usando matrices simétricas, que hemos estudiado.

El punto clave vuelve a ser si las matrices simétricas se diagonalizan en una base ortonormal adecuada. Esto resulta ser cierto y es el teorema profundo o la parte técnica medular del programa, pues el resto de los lemas e ideas que usamos se generalizan facilmente. Así que diagonalizando, cambiando escala, permutando variables y cambiando signo si es necesario, cualquier matriz es equivalente a una matriz diagonal con un bloque de 1 al principio, digamos que i de ellos, luego —si acaso— un bloque de -1, y finalmente —puede ser— un bloque de 0.

$$i\left\{\vphantom{\begin{matrix}1\\0\\1\end{matrix}}\right. \begin{pmatrix} 1 & 0 & & & \cdots & & & 0 \\ 0 & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & -1 & & & & \\ \vdots & & & & \ddots & & & \vdots \\ & & & & & -1 & & \\ & & & & & & 0 & \\ & & & & & & & \ddots \\ 0 & & & & \cdots & & & 0 \end{pmatrix} \left.\vphantom{\begin{matrix}1\\0\\1\\1\\1\\1\\1\\1\\1\end{matrix}}\right\} n+1$$

Si tiene ceros se le llama *degenerada*, pues esencialmente viene de una dimensión anterior (el polinomio homogéneo se expresa, en la base adecuada, con menos de las $n+1$ variables).

Las no degeneradas son las interesantes, y para éstas —como sólo nos interesan sus ceros— podemos suponer que hay a lo más tantos -1 como 1, es decir que $i \geq (n+1)/2$. Al número de -1, *i.e.* a $(n+1-i)$, se le llama el *índice* de la forma cuadrática no degenerada. Las de índice 0 definen el conjunto vacío tanto proyectiva como afinmente. Las de índice 1, aunque son todas equivalentes proyectivamente, dan lugar a tres tipos de cuádricas afines: los elipsoides (cuyo representante canónico es la esfera de dimensión n), los paraboloides que son tangentes al hiperplano al infinito y los hiperboloides de dos hojas que son cortados, separados, por el hiperplano al infinito que, a su vez, las intersecta en una cuádrica de índice 1 en una dimensión menos. Y para cada índice (que llega hasta $(n+1)/2$ para n impar y hasta $n/2$ para n par) la correspondiente cuádrica proyectiva tiene tantas manifestaciones afines como haya tipos de hiperplanos respecto a ella, que son los que se pueden mandar al infinito.

Capítulo 8

Geometría hiperbólica

Como ya hemos mencionado, la geometría hiperbólica se desarrolló de manera abstracta en el estudio del Quinto Postulado, pues surge al negarlo pluralizando: *son muchas las paralelas a una recta que pasan por un punto fuera de ella.* Haciendo esta suposición, a principios del siglo XIX unos cuantos matemáticos como Lobachevsky y Bólyai sentaron sus fundamentos. Estaban convencidos de su solidez lógica, y por tanto de su "existencia o posibilidad", pero estaban también en franca minoría. Se les tachaba de extravagancia o vil locura entre la comunidad científica. Se cuenta incluso que matemáticos de renombre, como el *"Principe de las Matemáticas"* Gauss, tenían resultados en geometría hiperbólica pero que no los hicieron públicos por temor al descrédito. Y no fue sino hasta que se descubrieron los modelos analíticos del plano hiperbólico cuando la comunidad matemática acabó por reconocerla. En la segunda mitad de ese siglo surgieron modelos del plano hiperbólico basados en la geometría euclidiana (en dimensión mayor). Fue entonces evidente su existencia; si la geometría euclidiana era consistente, también lo sería la hiperbólica.

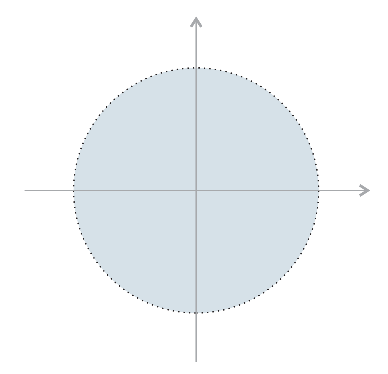
Nuestra presentación será analítica, y no axiomática, pero que quede este párrafo introductorio como homenaje a sus precursores que se dieron el lujo de tomarse la libertad de creer lo increíble, de crearlo y demostrarlo; sufrieron por ello, pero el tiempo les dio la razón.

8.1 El plano hiperbólico

Para definir la geometría hiperbólica (en dimensión 2), daremos muy fácilmente sus *puntos* (que forman el plano hiperbólico $\mathbb{H}^2$); sus subconjuntos distinguidos llamados *líneas*, y su *grupo de transformaciones hiperbólicas.* Lo que nos costará mucho más trabajo es demostrar que se merece el rimbombante nombre de "*Geometría*", que es homogéneo (igual en todos lados y en todas las direcciones) y que tiene nociones de distancia y ángulo de tal manera que el grupo sea justo el de isometrías.

8.1.1 Puntos

El *plano hiperbólico*, $\mathbb{H}^2$, consiste de los puntos interiores del círculo unitario, pero pensando $\mathbb{R}^2$ como la carta coordenada canónica del plano proyectivo $\mathbb{P}^2$, es decir,



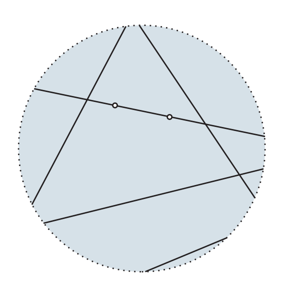
$$\begin{aligned}\mathbb{H}^2 := \quad & \left\{(x,y) \in \mathbb{R}^2 \mid x^2+y^2<1\right\} \\ \cong \quad & \left\{[x:y:z] \in \mathbb{P}^2 \mid x^2+y^2-z^2<0\right\}.\end{aligned}$$

Obsérvese que el círculo unitario $\mathbb{S}^1$ no es parte de $\mathbb{H}^2$, pero que está ligado a él, es su frontera, y a veces nos referiremos a $\mathbb{S}^1$ como "*el círculo al infinito*" del plano hiperbólico, y a sus puntos, como *puntos al infinito*, aunque, insistimos, los puntos al infinito del plano hiperbólico **no** son parte del plano hiperbólico (no hay que confundirse con el caso proyectivo).

A esta manera de ver al plano hiperbólico se le conoce como el *modelo proyectivo* o *modelo de Klein*(hay otros modelos más que veremos adelante) y es importante pensarlo como subconjunto del plano proyectivo. Que además $\mathbb{H}^2$ haya quedado dentro de la carta canónica, $\mathbb{R}^2$, y justo dentro de su círculo unitario, es una simple convención para hacer más sencillos los cálculos, pero no es una condición esencial como veremos en el capítulo siguiente; pudimos haber tomado el interior de cualquier cónica.

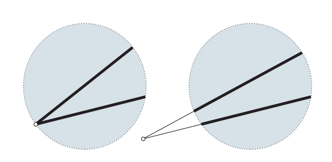
8.1.2 Líneas

Las *líneas*, o *rectas*, del plano hiperbólico son sus intersecciones con líneas proyectivas (o euclidianas, da lo mismo, pues $\mathbb{H}^2$ está en una región acotada). Es decir, son segmentos abiertos cuyos *extremos* están en $\mathbb{S}^1$ (son puntos al infinito hiperbólico), y corresponden a las *cuerdas* de $\mathbb{S}^1$.

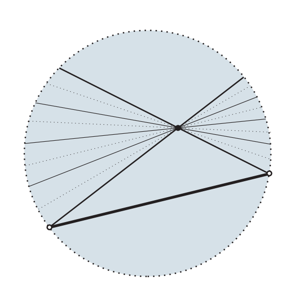


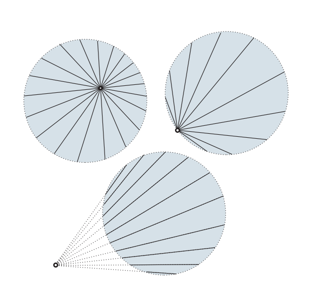
Es claro entonces que por cada par de puntos en $\mathbb{H}^2$ hay una única línea que pasa por ellos: la intersección con la correspondiente proyectiva (y por tanto, que si dos líneas en el plano hiperbólico se intersectan, entonces se intersectan en un único punto). Pero además también es clarísimo que se cumple la negación plural del Quinto. Por un punto fuera de una línea pasan muchas líneas que no la intersectan; veremos que son "tantas como un segmento cerrado".

Para ser más precisos, se pueden distinguir dos tipos de no intersección. Se dice que dos líneas hiperbólicas que no se intersectan son *paralelas* si la intersección de sus correspondientes líneas proyectivas es un punto al infinito (hiperbólico), es decir, si está en $\mathbb{S}^1$; y que son *ultraparalelas* si es un punto exterior. La otra posibilidad para dos líneas hiperbólicas es que sean *concurrentes* y se intersecten en $\mathbb{H}^2$.



Ahora sí, más precisamente, por un punto $\mathbf{p}$ fuera de una línea ℓ pasan dos paralelas que parten las rectas por $\mathbf{p}$ en dos sectores (o segmentos), las concurrentes con ℓ y las ultraparalelas con ℓ.





Tenemos entonces tres tipos de *haces* de líneas en el plano hiperbólico: los *concurrentes*, los *paralelos* y los *ultraparelelos*, que consisten en tomar las rectas hiperbólicas de haces proyectivos concurrentes; lo que los distingue es en dónde queda el punto de concurrencia.

Vale la pena observar que en el plano proyectivo sólo hay un tipo de haz, el concurrente, y que en el plano euclidiano hay dos tipos, el concurrente y el paralelo.

8.1.3 Transformaciones

El *grupo de transformaciones hiperbólicas* consiste de todas las transformaciones proyectivas que dejan en su lugar al plano hiperbólico, y por lo tanto a su frontera,

$$\begin{aligned} \mathrm{Hip}(2) : &= \left\{ \mathsf{f} \in \mathrm{Pr}(2) \mid \mathsf{f}\left(\mathbb{H}^2\right) = \mathbb{H}^2 \right\} \\ &= \left\{ \mathsf{f} \in \mathrm{Pr}(2) \mid \mathsf{f}\left(\mathbb{S}^1\right) = \mathbb{S}^1 \right\} \ ; \end{aligned}$$

es, por así decirlo, el grupo de "simetrías proyectivas" del círculo unitario.

Por el momento no es relevante que la igualdad de estos dos conjuntos esté bien demostrada –aunque lo argumentaremos intuitivamente a continuación–, pues usaremos como caballito de batalla otra definición con un enfoque técnico más analítico. Lo importante por ahora es que las transformaciones hiperbólicas sean naturales y fáciles de definir. Son lo que las transformaciones proyectivas, que en principio ya conocemos, le hacen al plano hiperbólico.

Que da lo mismo fijar el círculo unitario que su interior (la igualdad previa de conjuntos), es consecuencia de que $\mathbb{S}^1$ parte el plano proyectivo, $\mathbb{P}^2$, en dos pedazos topológicamente distintos: el interior que es un disco abierto (y que hemos llamado $\mathbb{H}^2$), y el exterior que es una banda de Moebius abierta cuya ánima o corazón es la recta al infinito. Basta entonces pedir a una transformación proyectiva que mande al círculo unitario en sí mismo, pues entonces tiene que mandar al interior, $\mathbb{H}^2$, en sí mismo y también al exterior en sí mismo; no los puede intercambiar.[1]

Ejemplos de transformaciones hiperbólicas son las rotaciones en el origen, pues mandan a $\mathbb{H}^2$ en él mismo. Pero podemos decir aún más, que

$$\mathrm{O}(2) \subset \mathrm{Hip}(2) \ ,$$

[1] Afirmaciones como ésta, que el disco es diferente de la banda de Moebius, aunque intuitivamente claras, sólo se pueden demostrar usando topología algebraica que se desarrolló en el siglo XX.

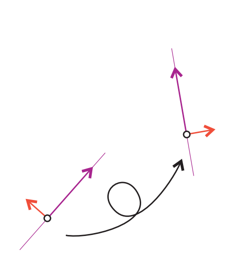
pues también las reflexiones en rectas por el origen dejan en su lugar al círculo unitario $\mathbb{S}^1$ (y a su interior $\mathbb{H}^2$). Demostraremos más adelante que las únicas transformaciones hiperbólicas que dejan fijo el origen son éstas, es decir que

$$\mathrm{O}(2) = \{f \in \mathrm{Hip}(2) \mid f(0) = 0\}. \tag{8.1}$$

Nótese que ésta es la misma propiedad que cumple el plano euclidiano $\mathbb{E}^2$: $\mathrm{O}(2)$ es el subgrupo de las isometrías que dejan fijo el origen. Y también es lo que cumplen el plano esférico (proyectivo rígido) o $\mathbb{S}^2$: las isometrías que dejan fijo un punto son rotaciones en él y reflexiones en una recta por él. Así que las tres geometrías son iguales local o infinitesimalmente.

La gran diferencia de enfoque es que en las geometrías anteriores (euclidiana y esférica) teníamos la métrica (nociones de distancia y ángulo) y de ahí dedujimos el grupo; ahora tenemos el grupo y queremos deducir que hay nociones de distancia y ángulo que lo hacen ser **el** grupo de isometrías.

Cuando logremos demostrar (8.1) y además que hay transformaciones hiperbólicas que mueven el origen (y por tanto que éste es un punto como cualquier otro), tendremos que $\mathbb{H}^2$ se comporta como las otras geometrías; que una transformación está determinada por lo que le hace a un punto, a una línea orientada por él y a uno de los lados que ésta define (si preserva o no orientación); una vez determinados estos valores, la métrica, o rigidez del plano, obliga al resto de la transformación (obsérvese que si el punto se queda fijo, los otros dos datos parametrizan $\mathrm{O}(2)$). La *homogeneidad* consiste en que todas estas ternas son equivalentes, "que se ve igual desde cualquier punto y en cualquier dirección". Primero estableceremos que éste es el caso de las transformaciones hiperbólicas, y entonces veremos que las nociones de distancia y ángulo están obligadas por el grupo de transformaciones. Pero para lograrlo necesitaremos más herramientas.



8.2 El espacio de Lorentz-Minkowski

Hemos definido al plano hiperbólico, $\mathbb{H}^2$, como el interior del círculo unitario, $\mathbb{S}^1$. Este último (en $\mathbb{R}^2$) está determinado por el polinomio x^2+y^2-1, y al considerarlo dentro del plano proyectivo, $\mathbb{P}^2$, su forma cuadrática asociada es el polinomio homogéneo en tres variables:

$$x^2 + y^2 - z^2,$$

que hemos llamado la *forma cuadrática de Lorentz*. Su matriz simétrica asociada es la *matriz de Lorentz*

$$L := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

y entonces podemos escribir

$$x^2 + y^2 - z^2 = (x, y, z) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{x}^\top L \mathbf{x}. \tag{8.2}$$

Una observación muy simple sobre esta última expresión ($\mathbf{x}^\top L \mathbf{x}$) nos brindará nuestra herramienta técnica fundamental. La observación es que el vector variable $\mathbf{x} \in \mathbb{R}^3$ aparece en dos lugares aislados de la expresión, y por tanto se puede extender a una expresión con dos variables vectoriales.

Para $\mathbf{x}$ y $\mathbf{y}$ en $\mathbb{R}^3$, definimos su *producto interno de Minkowski* como el número real

$$\begin{aligned} \mathbf{x} \bullet_L \mathbf{y} := & \quad \mathbf{x}^\top L \mathbf{y} \\ = & \quad \mathbf{x} \cdot (L\mathbf{y}) = (L\mathbf{x}) \cdot \mathbf{y}. \end{aligned} \tag{8.3}$$

Recordemos que tanto en la geometría euclidiana (capítulo **??**) como en la esférica (capítulo **??**), las nociones métricas básicas de distancia y ángulo se definieron usando como herramienta el producto interior. Lo que el producto interior es a la forma cuadrática euclidiana ($x^2 + y^2 + z^2$) —algo más general que, en particular, la define—, lo será ahora el producto interno de Minkowski respecto a la forma cuadrática de Lorentz (8.2); y de manera análoga a como se usó el producto interior euclidiano en las otras geometrías, el de Minkowski será fundamental para definir las nociones geométricas básicas en el plano hiperbólico y entender sus transformaciones.

En coordenadas, si $\mathbf{x} = (x_1, x_2, x_3)^\top$ y $\mathbf{y} = (y_1, y_2, y_3)^\top$, el producto interno de Minkowski toma la expresión

$$\begin{aligned} \mathbf{x} \bullet_L \mathbf{y} &= (x_1, x_2, x_3) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \\ &= x_1 y_1 + x_2 y_2 - x_3 y_3 \ , \end{aligned}$$

y es una función de $\mathbb{R}^3 \times \mathbb{R}^3$ en $\mathbb{R}$, que claramente cumple (ver siguiente ejercicio):

$$\begin{aligned} \mathbf{x} \bullet_L \mathbf{y} &= \mathbf{y} \bullet_L \mathbf{x} && \text{(es \textit{simétrica})}, \\ (\alpha \mathbf{x} + \beta \mathbf{y}) \bullet_L \mathbf{z} &= \alpha (\mathbf{x} \bullet_L \mathbf{z}) + \beta (\mathbf{y} \bullet_L \mathbf{z}) && \text{(es \textit{lineal})}. \end{aligned} \tag{8.4}$$

La segunda propiedad es que la función es lineal en la primera variable al fijar la segunda. Junto con la simetría, implica que así mismo es lineal en la segunda variable y entonces se dice que es *bilineal*. Entonces también se le llama *la forma bilineal de Minkowski* o de *Lorentz* o de *Lorentz-Minkowski*; en breve, comentaremos sobre el porqué de estos personajes asociados.

EJERCICIO 8.1 Demuestra las ecuaciones (8.4) para cualquier $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3$ y $\alpha, \beta \in \mathbb{R}$.

8.2.1 L-norma

Así como la norma (euclidiana, vamos a especificar) se obtiene del producto interior, podemos pensar la forma cuadrática de Lorentz como una "norma". De la fórmula clásica $|\mathbf{x}|^2 = \mathbf{x}\cdot\mathbf{x}$, vamos a piratear la notación para la *norma de Lorentz-Minkowski*, o L-*norma*, de un vector $\mathbf{x} \in \mathbb{R}^3$:

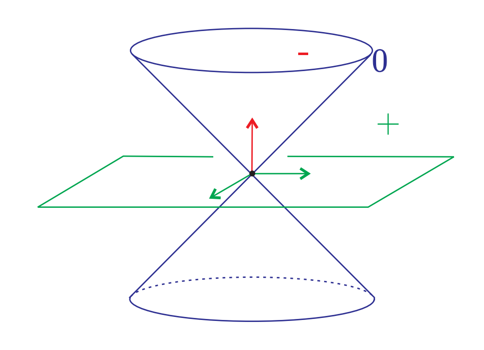
$$|\mathbf{x}|_L^2 := \mathbf{x} \bullet_L \mathbf{x} = \mathbf{x}^\top L\, \mathbf{x}$$

(que en coordenadas se escribe $x^2+y^2-z^2$ (8.2)), donde hay que pensar el "cuadrado", *i.e.*, el exponente 2, simplemente como notación. Pero notación afortunada pues claramente se cumple que

$$|t\,\mathbf{x}|_L^2 = t^2\ |\mathbf{x}|_L^2\,,$$

de tal manera que, como ya lo habíamos remarcado para cualquier forma cuadrática, el signo se preserva en las rectas por el origen.

La gran diferencia con el caso clásico es que ahora tenemos vectores de norma negativa o cero. Démosles nombres:



$$\text{Un vector } \ \mathbf{x} \in \mathbb{R}^3 \ \text{ o su recta } \ [\mathbf{x}] \quad \text{es} \quad \begin{cases} \textit{espacial,} & \text{si} \quad |\mathbf{x}|_L^2 > 0\ ; \\ \textit{temporal,} & \text{si} \quad |\mathbf{x}|_L^2 < 0\ ; \\ \textit{luz,} & \text{si} \quad |\mathbf{x}|_L^2 = 0\ . \end{cases}$$

Geométricamente, los vectores luz forman el cono sobre $\mathbb{S}^1$ (en el plano $z = 1$); los temporales están dentro del cono, y los espaciales fuera de él. Así, por ejemplo, los vectores canónicos $\mathbf{e}_1$ y $\mathbf{e}_2$ son espaciales, como cualquier otro vector en el plano xy; mientras que el eje z es temporal y las líneas luz son las que tienen un ángulo de 45° respecto al plano xy (y con el eje z).

Nota histórica

La nomenclatura que estamos usando viene de la Teoría de la Relatividad Especial que se desarrolló a principios del siglo XX. Albert Einstein es el personaje principal con quien se asocia su concepción, pero muchos otros científicos contribuyeron. Entre estos destacan como precursores el físico holandés Hendrik A. Lorentz y el matemático francés Henri Poincaré. Unos años despues del famoso trabajo de Einstein de 1905, el matemático alemán Hermann Minkowski reformuló en términos geométricos la Teoría de la Relatividad Especial.[2] Puesto que nuestra motivación y línea de razonamiento no es la que originalmente tenían ellos, es difícil establecer aquí con precisión qué es lo que históricamente se debe a cada cuál. Nuestra asociación de nombres (Lorentz y/o

[2] Un buen recuento de esta historia puede verse en [1].

Minkowski) a los objetos matemáticos que han aparecido no debe entonces tomarse en forma muy literal; más bien es un homenaje y un intento de apegarse a la terminología usada en la literatura (que no es única). Sin embargo, vale la pena tratar de precisar algo sobre el uso de los términos "espacial", "temporal" y "luz" (o quizá debía ser "lumínica").

El problema físico básico que da lugar a la Relatividad Especial tiene que ver con cómo se relacionan distintos sistemas de referencia. Para distintos observadores, un mismo evento o sucesión de eventos físicos pueden parecer diferentes; pero debería haber una "realidad física única". Se resuelve el dilema incluyendo el parámetro "tiempo" dentro de un continuo espacio-tiempo que permita hacer mediciones a observadores con respecto a "sus" coordenadas de tiempo y espacio, abandonando la idea de que éstas son absolutas: se relativizan. En nuestro caso ($\mathbb{R}^3$ con el producto interno de Minkowski), las dos primeras coordenadas, x y y, son las coordenadas espaciales y la tercera, z, es la coordenada temporal (en los textos con motivación estrictamente física se usaría una t en vez de nuestra z). Es un modelo de un universo de dos dimensiones espaciales (uno más parecido al que habitamos usaría una tercera coordenada espacial, z digamos, y una cuarta, ahora sí t, considerando entonces la forma cuadrática $x^2 + y^2 + z^2 - t^2$). A un punto en él se le llama un *evento* (un lugar en el espacio en un tiempo dado), y una partícula, o un observador, es una curva continua estrictamente ascendente, una *curva de vida*, digamos.

Parece una contradicción pues estamos dando coordenadas absolutas, pero sólo lo son para el evento 0, en el origen, y suponerlo como observador estacionario es determinar que su curva de vida es el eje z. Un observador inercial (en movimiento inercial uniforme) que coincida con el evento 0 es justo una recta por el origen. Que nada, excepto los fotones, pueda viajar a la velocidad de la luz (normalizada para que sea 1), significa que las líneas que hemos llamado luz (las del cono) son las trayectorias de los fotones que emanan de 0 y que las posibles líneas de vida inerciales del evento 0 son lo que hemos llamado líneas temporales (dentro del cono). Un evento espacial es, entonces, un evento sin relación posible con el 0, pero para un evento temporal hay la posibilidad de comunicación (por mensajes de luz, digamos).

Hasta aquí, y así de vago, medio se explican los términos o vocablos que estamos usando. Lo maravilloso es que las mismas herramientas matemáticas, y por tanto terminologías muy similares, sirvan para explicar y estudiar fenómenos aparentemente tan lejanos. Por un lado, que la velocidad de la luz es una constante irrebasable (el famoso experimento de Michelson y Morley) y su devenir en la Teoría Especial de la Relatividad, y por el otro la existencia de la geometría hiperbólica. En ambas, el producto interno de Minkowski es importante; por ejemplo, la noción de ortogonalidad que se desprende de él da, por un lado, el *plano de simultaneidad* para un observador inercial y para nosotros dará el concepto fundamental de polaridad.[3]

[3] Para un enfoque más físico de la geometría en el espacio de Minkowski ver [16] y [17].

8.2.2 L-ortogonalidad y polaridad

Lo primero que surge naturalmente del producto interior, o que por ahí se va con la norma, es la noción euclidiana de ortogonalidad. De igual manera, de la forma bilineal de Lorentz-Minkowski podemos sacar un análogo que tendrá un importante significado geométrico.

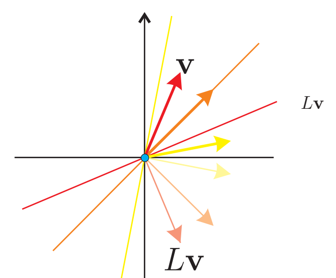
Dos vectores $\mathbf{v}$ y $\mathbf{u}$ en $\mathbb{R}^3$ son L-*ortogonales* si

$$\mathbf{v} \bullet_{\mathsf{L}} \mathbf{u} = 0.$$

Y podemos definir el *plano* L-*ortogonal* a $\mathbf{v} \in \mathbb{R}^3$, $\mathbf{v} \neq \mathbf{0}$, como

$$\Pi_{\mathsf{L}\mathbf{v}} := \left\{ \mathbf{x} \in \mathbb{R}^3 \mid \mathbf{v} \bullet_{\mathsf{L}} \mathbf{x} = 0 \right\}.$$

Veamos que nuestra definición de "plano L-ortogonal" es efectivamente un plano en $\mathbb{R}^3$, es decir, que estamos usando el vocablo "plano" de manera adecuada. Puesto que L como matriz es no singular -de hecho $\mathsf{L}^{-1} = \mathsf{L}$ pues L como transformación lineal es la reflexión en el plano xy-, se tiene que



$$\mathbf{v} \neq \mathbf{0} \quad \Leftrightarrow \quad \mathsf{L}\mathbf{v} \neq \mathbf{0}.$$

Entonces, como $\mathbf{v} \bullet_{\mathsf{L}} \mathbf{x} = (\mathsf{L}\mathbf{v}) \cdot \mathbf{x}$, donde " $\cdot$ " sigue siendo el producto interior euclidiano, tenemos que $\Pi_{\mathsf{L}\mathbf{v}}$ es el plano (euclidianamente) ortogonal a $\mathsf{L}\mathbf{v}$ que es, efectivamente, un plano pues $\mathsf{L}\mathbf{v} \neq \mathbf{0}$. De esta demostración podemos extraer, además del lema que viene, que si $\mathbf{v}$ gira de la vertical hacia el plano horizontal xy, su plano L-ortogonal gira del horizontal al vertical pero en sentido contrario, pues es el euclidianamente ortogonal a su reflejado en el plano xy.

Lema 8.1 *Cualquier plano por el origen en* $\mathbb{R}^3$ *es el plano* L-*ortogonal de algún* $\mathbf{v} \in \mathbb{R}^3$, $\mathbf{v} \neq \mathbf{0}$. *Además,* $\Pi_{\mathsf{L}\mathbf{v}} = \Pi_{\mathsf{L}\mathbf{u}}$ *si y sólo si* $[\mathbf{v}] = [\mathbf{u}]$.

Demostración. Sabemos que cualquier plano por el origen Π es euclidianamente ortogonal a algún $\mathbf{v}' \in \mathbb{R}^3$. Sea $\mathbf{v} = \mathsf{L}^{-1}\mathbf{v}' = \mathsf{L}\mathbf{v}'$, entonces $\Pi = \Pi_{\mathsf{L}\mathbf{v}}$. Además, $\Pi_{\mathsf{L}\mathbf{v}} = \Pi_{\mathsf{L}\mathbf{u}}$ si y sólo si $[\mathsf{L}\mathbf{v}] = [\mathsf{L}\mathbf{u}]$ y, como L es no singular, esto sucede si y sólo si $[\mathbf{v}] = [\mathbf{u}]$. □

Con este lema, podemos regresar al plano proyectivo, $\mathbb{P}^2$, pues para un punto $p = [\mathbf{v}]$ en él, tenemos definida una línea (la correspondiente al plano $\Pi_{\mathsf{L}\mathbf{v}}$) que llamaremos la *línea polar* de p, y denotaremos $p^{\perp}$ (léase "p-polar"); es decir,

$$p^{\perp} = [\mathbf{v}]^{\perp} := \ell_{[\mathsf{L}\mathbf{v}]} = \left\{ [\mathbf{x}] \in \mathbb{P}^2 \mid \mathbf{v} \bullet_{\mathsf{L}} \mathbf{x} = 0 \right\},$$

que, como veremos a continuación, es la línea polar respecto al círculo unitario que habíamos definido en el capítulo 2 y que entonces habíamos denotado ℓ_p. Pero además,

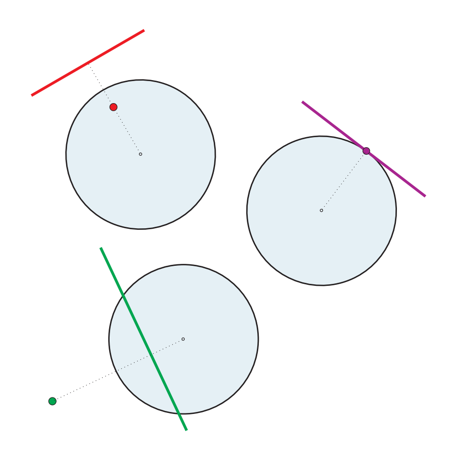
el lema implica que para cualquier línea proyectiva ℓ existe un punto bien definido, que llamaremos su *punto polar*, denotado $\ell^{\perp}$, del cual ℓ es su línea polar $\left(\left(\ell^{\perp}\right)^{\perp} = \ell\right)$. Veamos cómo es esta correspondencia en términos geométricos

Si tomamos un punto $p = (a, b) \in \mathbb{R}^2 \subset \mathbb{P}^2$, entonces su vector asociado en $\mathbb{R}^3$ es $\mathbf{v} = (a, b, 1)$, cuya línea polar, por definición, tiene como polinomio homogéneo asociado a

$$\mathbf{v} \bullet_L \mathbf{x} = (a, b, 1) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = ax + by - z.$$

Que nos da, de regreso en $\mathbb{R}^2$, haciendo $z = 1$, que la línea polar de $p = (a, b)$ es

$$p^{\perp} : \quad ax + by = 1$$



que, para $(a, b) \neq (0, 0)$, es la línea ortogonal al vector (a, b) y cuya distancia al origen es el inverso de la norma de (a, b); por ejemplo, para el punto $(a, 0)$ es la recta $x = 1/a$. Tenemos entonces tres casos: si $p \in \mathbb{H}^2$ (si p es *temporal*, podríamos decir) entonces $p^{\perp}$ es una línea afuera del círculo unitario (es una línea *espacial*); si $p \in \mathbb{S}^1$ (es *luz*) entonces su polar $p^{\perp}$ es la tangente a $\mathbb{S}^1$ en él (y se tiene que $p \in p^{\perp}$ pues $a^2 + b^2 = 1$), y si p es *espacial* (es exterior al círculo) entonces $p^{\perp}$ es una línea hiperbólica (pasa por $\mathbb{H}^2$).

En el párrafo anterior (que corresponde a lo hecho en el capítulo 2) nos faltaron ciertos casos proyectivos que analizar. Primero, la línea polar del origen (el centro del círculo, $(0, 0)$) es precisamente la línea al infinito, definida por el polinomio homogéneo $-z$. Y por último, para los puntos al infinito, con vector asociado $(a, b, 0)$, se tiene que sus líneas polares son las líneas por el origen perpendiculares a la dirección correspondiente, con ecuación $ax + by = 0$.

Tenemos entonces, y en correspondencia al lema anterior, que la polaridad es una biyección entre puntos y líneas proyectivas. Además cumple el *principio de dualidad*:

$$p \in q^{\perp} \quad \Leftrightarrow \quad q \in p^{\perp}, \tag{8.5}$$

pues si $p = [\mathbf{v}]$ y $q = [\mathbf{u}]$, ambas condiciones son, por definición, que $\mathbf{v} \bullet_L \mathbf{u} = 0$; y en este caso podemos decir entonces que los puntos p y q son (L-) *ortogonales*. Y más aún, podemos definir que dos rectas η y ξ (usaremos las letras griegas η "eta", ξ "xi" y ζ "dseta" para designar líneas proyectivas) son *ortogonales* si sus puntos polares lo son, es decir, si $\eta^{\perp} \in \xi$ (que es equivalente a $\xi^{\perp} \in \eta$). El caso más importante es cuando η y ξ son líneas hiperbólicas concurrentes, *i.e.*, cuando $\eta \cap \xi \in \mathbb{H}^2$, pues que sean ortogonales va a querer decir que su ángulo hiperbólico es de $90°$; ya es una noción métrica.

Hemos definido la polaridad de manera algebraica, pero tiene un fuerte contenido geométrico puesto que la línea polar de un punto en $\mathbb{S}^1$ es su línea tangente y se cumple el principio de dualidad (8.5). Así, si η es una línea hiperbólica y q_1 y q_2 son sus puntos al infinito, es decir, si $\{q_1, q_2\} = \eta \cap \mathbb{S}^1$, entonces la intersección de las tangentes a $\mathbb{S}^1$ en q_1 y q_2 es el punto polar de η, es decir,

$$\eta^\perp = q_1^\perp \cap q_2^\perp.$$

Pues si declaramos $p = q_1^\perp \cap q_2^\perp$ entonces $q_1 \in p^\perp$ y $q_2 \in p^\perp$ por (8.5), entonces $p^\perp$ tiene que ser la recta por q_1 y q_2 que es η, y por tanto $\eta^\perp = \left(p^\perp\right)^\perp = p$. E inversamente, tenemos que si p es un punto espacial, su polar es la línea hiperbólica que pasa por sus *pies* en $\mathbb{S}^1$, es decir, los dos puntos en $\mathbb{S}^1$ cuyas tangentes pasan por p.

EJERCICIO 8.2 Calcula los pies del punto $(2, 0)$ en el círculo unitario encontrando la ecuación de su línea polar y sus intersecciones con $\mathbb{S}^1$. ¿Puedes dar una fórmula para los pies del punto (a, b) (con $a^2 + b^2 > 1$, por supuesto)?

EJERCICIO 8.3 ¿Cómo construyes con regla y compás la línea polar de un punto hiperbólico? (Supón que el compás te sirve para trazar tangentes a un círculo.)

EJERCICIO 8.4 Demuestra que dos rectas hiperbólicas son concurrentes (se intersectan dentro de $\mathbb{H}^2$) si y sólo si la recta que pasa por sus polares es espacial.

EJERCICIO 8.5 Demuestra que las líneas polares de puntos colineales son concurrentes.

EJERCICIO 8.6 Demuestra que, dado un punto $p \in \mathbb{H}^2$ y una recta hiperbólica η, existe una única recta hiperbólica ξ ortogonal a η y que pasa por p. (Primero olvídate de las condiciones sobre dónde están los puntos y las líneas.)

EJERCICIO 8.7 En $\mathbb{H}^2$, demuestra que dos rectas ultraparalelas tienen una única recta ortogonal común.

EJERCICIO 8.8 Demuestra que los haces ultraparalelos están formados por todas las ortogonales a una recta dada.

EJERCICIO 8.9 ¿Cuáles son las rectas ortogonales a una recta *luz*, es decir, polar de un punto luz?

8.2.3 Ternas ortogonales

Tres puntos a, b, c en $\mathbb{P}^2$ forman una *terna ortogonal* si dos a dos son ortogonales. Como ejemplo obvio, tenemos la *terna canónica*, $[\boldsymbol{e}_1]$, $[\boldsymbol{e}_2]$, $[\boldsymbol{e}_3]$ pues $\boldsymbol{e}_i \bullet_L \boldsymbol{e}_j = 0$ para $i \neq j$. Obsérvese que si bien el producto interior de Minkowski no está bien definido (ni en signo) para puntos en $\mathbb{P}^2$, pues $(t\,\boldsymbol{u}) \bullet_L (s\,\boldsymbol{v}) = (t\,s)\,(\boldsymbol{u} \bullet_L \boldsymbol{v})$, y entonces depende de los representantes; la L-ortogonalidad sí lo está, pues si este producto es 0 para algún par de representantes, lo es para cualquiera.

Veremos ahora que existen muchas ternas ortogonales; tantas como líneas con un punto distinguido en el plano hiperbólico. Consideremos cualquier punto $\mathfrak{c} \in \mathbb{H}^2$; empezamos con el tercer punto por analogía con la terna canónica donde $[\boldsymbol{e}_3]$ es el punto hiperbólico. Tomemos ahora cualquier punto $\mathfrak{b} \in \mathfrak{c}^\perp$ (o, lo que es lo mismo, cualquier recta hiperbólica $\mathfrak{b}^\perp$ que pase por $\mathfrak{c}$). Y ahora sí, $\mathfrak{a}$ queda determinado por

$$\mathfrak{a} = \mathfrak{b}^\perp \cap \mathfrak{c}^\perp. \tag{8.6}$$

Está bien definido pues $\mathfrak{b}^\perp \neq \mathfrak{c}^\perp$ ya que la primera recta es hiperbólica (contiene a $\mathfrak{c}$ pues $\mathfrak{c} \in \mathfrak{b}^\perp$) y la segunda espacial, y además, por definición, $\mathfrak{a}$ es ortogonal a $\mathfrak{b}$ y a $\mathfrak{c}$.

Nótese que en una terna ortogonal, el punto polar a la línea por dos de ellos es el tercero; pues, *e.g.*, $\mathfrak{b} \in \mathfrak{a}^\perp$ y $\mathfrak{c} \in \mathfrak{a}^\perp$ implican que $\langle \mathfrak{b}\mathfrak{c} \rangle = \mathfrak{a}^\perp$, que es la ecuación dual de (8.6).

No es casualidad que en el ejemplo nos hayan salido dos puntos espaciales y uno temporal.

Lema 8.2 *Sea* $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ *una terna ortogonal; entonces dos de los puntos son espaciales y uno es temporal.*

Demostración. Observemos primero que si uno de los puntos es hiperbólico entonces, como su línea polar es espacial y en ella están los otros dos puntos, el lema se cumple; la construcción es como arriba.

Supongamos que alguno de los puntos, digamos $\mathfrak{a}$, es espacial. Pongamos las rectas por $\mathfrak{a}$ en biyección natural —por intersección— con los puntos de su línea polar $\mathfrak{a}^\perp$. Los dos pies de $\mathfrak{a}$ parten los puntos de $\mathfrak{a}^\perp$ en dos segmentos abiertos: puntos (y correspondientemente, líneas por $\mathfrak{a}$) hiperbólicos, y puntos (y líneas por $\mathfrak{a}$) espaciales. Sabemos que $\mathfrak{b}$ está en $\mathfrak{a}^\perp$ y que

$$\mathfrak{c} = \mathfrak{a}^\perp \cap \mathfrak{b}^\perp.$$

Si $\mathfrak{b}$ es espacial, entonces la línea $\mathfrak{b}^\perp$ es hiperbólica y por tanto también lo es $\mathfrak{c}$. Si $\mathfrak{b}$ es luz ($\mathfrak{b} \in \mathfrak{b}^\perp$), entonces $\mathfrak{c} = \mathfrak{a}^\perp \cap \mathfrak{b}^\perp = \mathfrak{b}$ y no tendríamos una terna. Nótese que esto implica que no hay puntos luz en la terna pues todos sus ortogonales (y distintos) son espaciales. Y hemos terminado. □

Podemos convenir entonces que una terna ortogonal $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ está ordenada y tiene como punto temporal al tercero, $\mathfrak{c}$. Y concluir que las **ternas ortogonales están en biyección natural con las parejas incidentes punto-recta** en el plano hiperbólico $\mathbb{H}^2$, pues una terna se define al escoger un punto $\mathfrak{c} \in \mathbb{H}^2$ junto con una recta $\mathfrak{b}^\perp$ que pasa por él; ya que entonces $\mathfrak{a}^\perp$ es la única ortogonal a $\mathfrak{b}^\perp$ por $\mathfrak{c}$ (ver Ejercicio 8.6), o bien se tiene que $\mathfrak{a} = \mathfrak{b}^\perp \cap \mathfrak{c}^\perp$.

EJERCICIO 8.10 Describe todas las ternas ortogonales que tienen el origen como punto temporal.

EJERCICIO 8.11 Encuentra una terna ortogonal que tenga al punto $c = [1 : 1 : 2]$ como punto temporal; haz un dibujo.

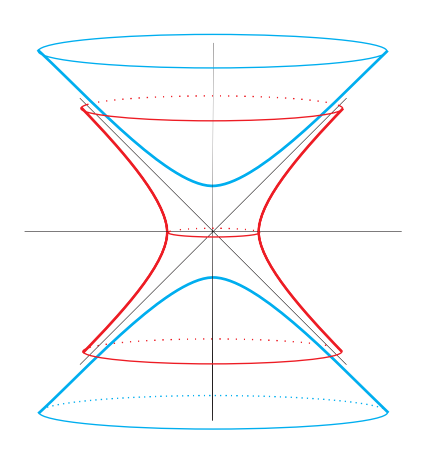
EJERCICIO 8.12 Describe geométricamente cómo construir con regla y compás la terna ortogonal que corresponde a un punto en el disco unitario y un segmento que pasa por él (el compás lo necesitas para trazar ortogonales euclidianas).

EJERCICIO 8.13 Sean a y b dos puntos espaciales. Demuestra que las líneas hiperbólicas $a^\perp$ y $b^\perp$ son concurrentes, paralelas o ultraparalelas correspondientemente a que la línea proyectiva por a y b, $\langle ab \rangle$, es espacial, lumínica (tangente a $\mathbb{S}^1$) o hiperbólica.

8.3 El grupo de transformaciones

Teniendo ya una multitud de ternas ortogonales en el plano proyectivo $\mathbb{P}^2$, las subiremos a $\mathbb{R}^3$ para construir ternas L-*ortonormales* y a partir de ellas entenderemos el grupo de transformaciones hiperbólicas. La analogía que hay que mantener en mente es que las matrices ortogonales (en el sentido euclidiano clásico) tienen como columnas a vectores que forman una base ortonormal y son justo las que preservan el producto interior.

8.3.1 Vectores L-unitarios y la cazuela hiperbólica

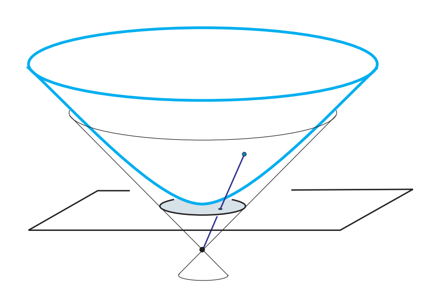


Decimos que un vector $x \in \mathbb{R}^3$ es L-*unitario* si su L-norma es 1 o -1, es decir, si en valor absoluto

$$\left| |x|_L^2 \right| = 1.$$

Tenemos entonces dos tipos de vectores unitarios: los temporales que forman un hiperboloide de dos hojas (el dado por la ecuación $x^2 + y^2 - z^2 = -1$), y los espaciales que forman un hiperboloide reglado de una hoja (dado por $x^2 + y^2 - z^2 = 1$). Esto se ve pues se obtienen, respectivamente, rotando en el eje z a las hipérbolas $y^2 - z^2 = -1$ y $y^2 - z^2 = 1$ que viven en el plano yz.

El cono de luz es el cono asintótico de ambos hiperboloides; por dentro se le unta el temporal de dos hojas y por fuera el espacial.



Notemos que el plano hiperbólico está en biyección natural con la hoja superior de los L-unitarios temporales, a la que llamaremos la *cazuela hiperbólica*. Pues si pensamos a $\mathbb{H}^2$ dentro del plano $z = 1$, en la carta coordenada canónica como habíamos quedado, la línea por el origen correspondiente a un punto en $\mathbb{H}^2$ es temporal e intersecta a ambas hojas de los vectores temporales unitarios, entonces le podemos asignar el punto que está en la cazuela hiperbólica y que tiene su tercera coordenada positiva. E inversamente, un punto en la cazuela hiperbólica representa un punto en el proyectivo que está en $\mathbb{H}^2$. Geometricamente es la proyección desde el origen.

A este modelo del plano hiperbólico (la cazuela) se le conoce a veces como "la esfera de radio imaginario" , pues son soluciones para la ecuación $|\mathbf{x}|_L^2 = -1$; pero nosotros le seguiremos llamando "la cazuela". Representa algo más realista respecto a la geometría hiperbólica en cuanto a distancias y ángulos se refiere, que el modelo proyectivo ($\mathbb{H}^2$ como subconjunto de $\mathbb{P}^2$) que hemos venido usando; pues, por lo menos, el infinito ya se ve bien lejos: no dan ganas de pegárselo. Además, las rectas hiperbólicas se ven en la cazuela como intersecciones con planos por el origen, que son una rama de una hipérbola.

Nuestros dos modelos de $\mathbb{H}^2$ seguirán siendo compatibles, pues la manera en que definiremos el grupo de transformaciones, a continuación, dejará fija, y actuará, también en la cazuela hiperbólica. Para no excedernos en notación, introduciendo un nuevo símbolo, abusaremos de ella **denotando a la cazuela hiperbólica también con** $\mathbb{H}^2$; el contexto nos dirá si nos referimos al modelo proyectivo (por ejemplo, si los puntos en $\mathbb{H}^2$ considerados son proyectivos), o bien si es la cazuela, cuando sus elementos son vectores en $\mathbb{R}^3$.

8.3.2 Bases L-ortonormales

Una terna ordenada de vectores $\mathbf{u}, \mathbf{v}, \mathbf{w}$ en $\mathbb{R}^3$ es una base L-*ortonormal* si los tres vectores son L-unitarios y dos a dos son L-ortogonales. Es claro que entonces la terna $[\mathbf{u}]$, $[\mathbf{v}]$, $[\mathbf{w}]$ en $\mathbb{P}^2$ es ortogonal, y por el Lema 8.2 uno de los vectores es temporal y los otros dos son espaciales. Supondremos siempre que el tercer vector, $\mathbf{w}$, es el que es temporal. A la matriz $B = (\mathbf{u}, \mathbf{v}, \mathbf{w})$ la llamaremos L-*ortogonal*, de tal manera que ternas L-ortonormales y matrices L-ortogonales se corresponden. (Como las bases ortonormales euclidianas y las matrices ortogonales.)

Obsérvese que la base canónica $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ es L-ortonormal (y la matriz identidad I es L-ortogonal).

La maravilla que cumplen las matrices L-ortogonales, como en el caso de las matrices ortogonales respecto al producto interior, es que al pensarlas como transformación lineal, dejan invariante la forma bilineal de Lorentz.

Teorema 8.3 *Sea* B *una matriz de* 3×3*; entonces son equivalentes:*

$$\begin{array}{rl} i) & \mathsf{B} \text{ es L-ortogonal;} \\ ii) & \mathbf{x} \bullet_{\mathsf{L}} \mathbf{y} = (\mathsf{B}\mathbf{x}) \bullet_{\mathsf{L}} (\mathsf{B}\mathbf{y}) \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^3; \\ iii) & \mathsf{B}^\top \mathsf{L} \mathsf{B} = \mathsf{L}. \end{array}$$

Demostración. Casi por n-ésima vez pero ahora con dos variables, hagamos el cálculo de

$$(\mathsf{B}\mathbf{x}) \bullet_{\mathsf{L}} (\mathsf{B}\mathbf{y}) = (\mathsf{B}\mathbf{x})^\top \mathsf{L}\,(\mathsf{B}\mathbf{y}) = \left(\mathbf{x}^\top \mathsf{B}^\top\right)(\mathsf{L}\mathsf{B})\,\mathbf{y} = \mathbf{x}^\top \left(\mathsf{B}^\top \mathsf{L} \mathsf{B}\right)\mathbf{y}.$$

Es claro entonces que (iii) $\Rightarrow$ (ii).

Expresemos la matriz $\mathsf{B}^\top \mathsf{L} \mathsf{B}$ en términos de las columnas de $\mathsf{B} = (\mathbf{u}, \mathbf{v}, \mathbf{w})$,

$$\begin{aligned} \mathsf{B}^\top \mathsf{L} \mathsf{B} &= \mathsf{B}^\top \mathsf{L}\,(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \mathsf{B}^\top\,(\mathsf{L}\mathbf{u}, \mathsf{L}\mathbf{v}, \mathsf{L}\mathbf{w}) \\ &= \begin{pmatrix} \mathbf{u}^\top \\ \mathbf{v}^\top \\ \mathbf{w}^\top \end{pmatrix} (\mathsf{L}\mathbf{u}, \mathsf{L}\mathbf{v}, \mathsf{L}\mathbf{w}) \\ &= \begin{pmatrix} \mathbf{u}^\top \mathsf{L}\mathbf{u} & \mathbf{u}^\top \mathsf{L}\mathbf{v} & \mathbf{u}^\top \mathsf{L}\mathbf{w} \\ \mathbf{v}^\top \mathsf{L}\mathbf{u} & \mathbf{v}^\top \mathsf{L}\mathbf{v} & \mathbf{v}^\top \mathsf{L}\mathbf{w} \\ \mathbf{w}^\top \mathsf{L}\mathbf{u} & \mathbf{w}^\top \mathsf{L}\mathbf{v} & \mathbf{w}^\top \mathsf{L}\mathbf{w} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{u} \bullet_{\mathsf{L}} \mathbf{u} & \mathbf{u} \bullet_{\mathsf{L}} \mathbf{v} & \mathbf{u} \bullet_{\mathsf{L}} \mathbf{w} \\ \mathbf{v} \bullet_{\mathsf{L}} \mathbf{u} & \mathbf{v} \bullet_{\mathsf{L}} \mathbf{v} & \mathbf{v} \bullet_{\mathsf{L}} \mathbf{w} \\ \mathbf{w} \bullet_{\mathsf{L}} \mathbf{u} & \mathbf{w} \bullet_{\mathsf{L}} \mathbf{v} & \mathbf{w} \bullet_{\mathsf{L}} \mathbf{w} \end{pmatrix}. \end{aligned}$$

Nótese que las seis condiciones para que $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ sean una base L-ortonormal (tres de normalidad y tres de ortogonalidad) son precisamente que las entradas de esta matriz simétrica sean justo las de la matriz de Lorentz, L (donde es importante nuestra convención de que $\mathbf{w}$ sea el temporal). Así que (iii) $\Leftrightarrow$ (i).

(ii) $\Rightarrow$ (iii). Puesto que $\mathbf{u} = \mathsf{B}\mathbf{e}_1$, $\mathbf{v} = \mathsf{B}\mathbf{e}_2$ y $\mathbf{w} = \mathsf{B}\mathbf{e}_3$, entonces si se cumple (ii) se tiene que esta última matriz es L, pues la base canónica es L-ortonormal. □

Más que por su demostración, que es muy simple, este teorema es tal por su fuerza. Nos da un subgrupo de GL (3) asociado a la forma bilineal de Lorentz (ver Ejercicio 8.14), que se denota $\mathsf{O}\,(2,1)$ pues se dice que la forma bilineal de Lorentz tiene *índice* $2, 1$ (por el número de "más" y "menos"). Generalizan estos grupos a los grupos ortogonales. Al preservar la forma bilineal de Lorentz, estas matrices preservan los "estratos" de su forma cuadrática. En particular manda vectores luz en vectores luz, temporales en temporales y a los espaciales en ellos mismos. Pero aún más, preservan las *superficies de nivel*, definidas por $|\mathbf{x}|_{\mathsf{L}}^2 = \mathrm{cte}$. Nuestro interes está en el subgrupo de las que preservan la cazuela hiperbólica, y esto es fácil de detectar pues la tercera columna, como vector, debe apuntar hacia arriba.

Sea

$$\mathcal{H}(2) := \left\{ B \in GL(3) \mid B^\top L B = L;\ \ B\boldsymbol{e}_3 \cdot \boldsymbol{e}_3 > 0 \right\}.$$

Estas matrices son las L-ortogonales que tienen su vector temporal unitario precisamente en la cazuela $\mathbb{H}^2$; las llamaremos *matrices hiperbólicas*. Forman un grupo (Ejercicio 8.15) de transformaciones lineales de $\mathbb{R}^3$ que mandan a $\mathbb{H}^2$ en sí mismo pues para $B \in \mathcal{H}(2)$

$$\mathbf{x} \in \mathbb{H}^2 \quad \Rightarrow \quad |\mathbf{x}|_L^2 = -1 \quad \Rightarrow \quad |B\mathbf{x}|_L^2 = B\mathbf{x} \bullet_L B\mathbf{x} = \mathbf{x} \bullet_L \mathbf{x} = |\mathbf{x}|_L^2 = -1;$$

y además $B\mathbf{x}$ tiene tercera coordenada positiva como $B\boldsymbol{e}_3$, pues el segmento de $\mathbf{x}$ a $\boldsymbol{e}_3$ es mandado al segmento de $B\mathbf{x}$ a $B\boldsymbol{e}_3$ y en éste la forma cuadrática de Lorentz es estrictamente negativa y por lo tanto no cruza el cono de luz.

Veremos que este grupo $\mathcal{H}(2)$ de *matrices hiperbólicas* se identifica naturalmente con el grupo Hip (2) de transformaciones hiperbólicas que definimos geométricamente en 8.1.3 (las transformaciones proyectivas que preservan el círulo unitario). Como una matriz en GL (3) define una transformación proyectiva, tenemos una función natural $\mathcal{H}(2) \to \text{Hip}(2)$, pues, como ya hemos dicho, las matrices L-ortogonales mandan al cono de luz ($\mathbb{S}^1$) y al conjunto de las líneas temporales ($\mathbb{H}^2$) en sí mismos. Esta función es una inclusión (inyectiva) pues dos matrices definen la misma transformación proyectiva sólo cuando son múltiplos; si dos matrices distintas L-ortogonales fueran múltiplos, la constante que las relaciona tendría que ser -1 pues sus vectores son L-unitarios, y entonces sólo una está en $\mathcal{H}(2)$, la que tiene entrada $3,3$ positiva. Así que (podemos pensar que)

$$\mathcal{H}(2) \subset \text{Hip}(2). \tag{8.7}$$

Antes de entrar de lleno a la prueba de la igualdad conviene ver algunas de las propiedades de las matrices hiperbólicas y algunos ejemplos.

EJERCICIO 8.14 Demuestra que para cualquier matriz $A \in GL(3)$ se tiene que $G_A := \left\{ B \mid B^\top A B = A \right\}$ es un subgrupo de GL(3); obsérvese que también hay que demostrar que los elementos de G_A son matrices invertibles. ¿Cuánto vale el determinante de estas matrices?

EJERCICIO 8.15 Demuestra que $\mathcal{H}(2)$ es un grupo.

EJERCICIO 8.16 Dada una matriz

$$B = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{pmatrix},$$

demuestra que $B^\top L B = L$ si y sólo si

$$B^{-1} = \begin{pmatrix} \alpha_1 & \beta_1 & -\gamma_1 \\ \alpha_2 & \beta_2 & -\gamma_2 \\ -\alpha_3 & -\beta_3 & \gamma_3 \end{pmatrix}.$$

EJERCICIO 8.17 Exhibe una transformación hiperbólica que mande al punto $[1/2, 1/2, 1]$ en el origen. (Usa el Ejercicio 8.11 y el anterior.)

8.3.3 Homogeneidad e isotropía

Veremos que todos los puntos de $\mathbb{H}^2$ son iguales (equivalentes según el grupo) —propiedad conocida como *homogeneidad*— y que, al fijar un punto, también son
1...Dibujos equivalentes todas las direcciones por él —*isotropía*—.

Consideremos una terna ortogonal $\mathsf{a}, \mathsf{b}, \mathsf{c}$ en $\mathbb{P}^2$. La podemos levantar a una base L-ortonormal como sigue. A c le corresponde un único vector $\boldsymbol{w}$ en la cazuela hiperbólica. A b (y respectivamente a a) le corresponde una recta por el origen en $\mathbb{R}^3$ que, por ser espacial ("acostadita"), intersecta al hiperboloide de L-unitarios espaciales justamente en dos puntos. A saber, si $\mathsf{b} = [\mathbf{b}]$ estos son

$$\pm\left(\sqrt{|\mathbf{b}|_L^2}\right)^{-1}\mathbf{b},$$

que tiene sentido pues $|\mathbf{b}|_L^2 > 0$. Llamemos $\boldsymbol{v}$ a cualquiera de ellos de tal manera que nuestras dos opciones son $\pm\boldsymbol{v}$. Y, respectivamente, sea $\mathbf{u} \in \mathbb{R}^3$ tal que $|\mathbf{u}|_L^2 = 1$ y $\mathsf{a} = [\mathbf{u}]$. Tenemos entonces que las cuatro matrices

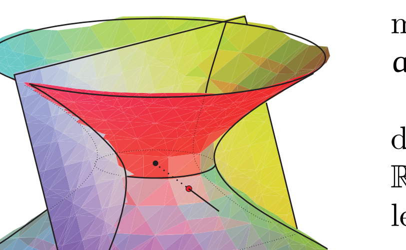
$$(\pm\mathbf{u}, \pm\boldsymbol{v}, \boldsymbol{w})$$

están en $\mathcal{H}(2)$ y mandan a la terna canónica $[\mathbf{e}_1], [\mathbf{e}_2], [\mathbf{e}_3]$ en la terna ortogonal $\mathsf{a}, \mathsf{b}, \mathsf{c}$ con que empezamos.

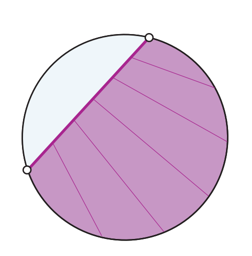
Consideremos el ejemplo más sencillo que es la terna canónica. Si tenemos que $(\mathsf{a}, \mathsf{b}, \mathsf{c}) = ([\mathbf{e}_1], [\mathbf{e}_2], [\mathbf{e}_3])$, las cuatro matrices a las que llegamos en el párrafo anterior son

$$\begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

que, efectivamente, dejan fijo al origen y a cada uno de los dos ejes (estamos pensando en la carta canónica, por supuesto), pero permutan los cuatro cuadrantes, les cambian orientación a los ejes. Forman además un grupito (el diédrico de orden dos, $\mathbf{D}_2$). Escoger entonces a alguna de las cuatro matrices tiene que ver con cómo se orientan las rectas $\mathsf{b}^\perp$ y $\mathsf{a}^\perp$. Veámoslo.

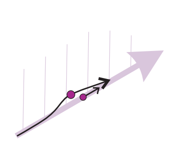


El vector $\boldsymbol{v}$ (seguimos pensando en el de arriba) no sólo define a la recta proyectiva $[\boldsymbol{v}]^\perp = \mathsf{b}^\perp$ sino también al plano en $\mathbb{R}^3$ que habíamos denotado con $\Pi_{L\boldsymbol{v}}$ $(: \boldsymbol{v} \bullet_L \mathbf{x} = 0)$, pero además le escoge un lado: en el que él está (bien definido, pues $\boldsymbol{v} \notin \Pi_{L\boldsymbol{v}}$ por ser espacial). De tal manera que la otra opción, $-\boldsymbol{v}$, escoge el lado opuesto de $\Pi_{L\boldsymbol{v}}$.

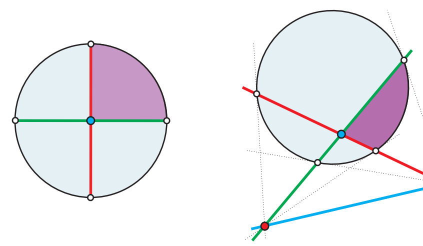


Definamos entonces una *línea hiperbólica orientada* como una línea hiperbólica junto con uno de sus dos lados en $\mathbb{H}^2$ distinguido. Y ahora sí, a una línea hiperbólica orientada podemos asociarle de manera continua y única un vector unitario espacial: el que está del mismo lado de su correspondiente plano por el origen que el lado distinguido de la recta (pensando que $\mathbb{H}^2$ está en la carta coordenada canónica contenida en $\mathbb{R}^3$).

Dos observaciones son pertinentes antes de seguir adelante. Primera, que nuestra nueva definición de orientación para rectas es equivalente a la —quizá más natural por ser intrínseca— de distinguir una dirección preferida de circulación para sus puntos. La equivalencia depende de adoptar una convención. *La regla de la autopista*, que —con perdón de los ingleses— reza: "siempre se rebasa por la izquierda". Así, si dos puntos viajan por una recta en su dirección intrínseca distinguida pero el que va atrás va más rápido, puede rebasar al lento por la izquierda distinguiendo entonces un lado; e inversamente, con un lado distinguido hay sólo una dirección intrínseca que lo usa para rebases civilizados. Sin embargo, esto no funciona para el plano proyectivo donde una recta no corta en dos "lados" pero sí se puede orientar intrínsecamente (el segundo rebase sería por el lado contrario), poniendo en duda nuestra nueva definición que involucra rectas proyectivas. Ésta es la segunda observación, junto que no hay problema. Pues sólo estamos hablando de líneas hiperbólicas que sí cortan al disco $\mathbb{H}^2$ en dos pedazos.



Por ejemplo, el eje de las x ($[\mathbf{e}_2]^\perp$) con su orientación positiva tiene como vector asociado a $\mathbf{e}_2$; y $\mathbf{e}_1$ tiene como recta hiperbólica orientada al lado derecho del eje y, dándole orientación intrínseca hacia abajo, ¡la negativa! Pero queda bien, pues si intersectamos los dos lados distinguidos, obtenemos el *primer cuadrante*. En general, a una base L-ortogonal $\mathbf{u}, \mathbf{v}, \mathbf{w}$ le podemos asociar un *cuadrante hiperbólico:* la intersección de los lados distinguidos de las dos rectas hiperbólicas ortogonales orientadas que determinan $\mathbf{u}$ y $\mathbf{v}$; su esquina es $[\mathbf{w}]$. (Obsérvese que es la imagen del primer cuadrante bajo la matriz L-ortonormal $B = (\mathbf{u}, \mathbf{v}, \mathbf{w})$.) Así que un cuadrante hiperbólico es una terna ortogonal con la información adicional de uno de sus cuatro cuadrantes distinguido; y la base canónica tiene asociado el primer cuadrante. Hemos demostrado, junto con el consabido argumento de "contagio grupal" —eso de que con una que vaya a todas entonces ya todas van contra todas—, el siguiente teorema.



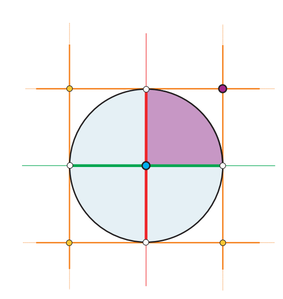
Teorema 8.4 *Las matrices hiperbólicas, $\mathcal{H}(2)$, están en correspondencia biyectiva natural —asociándoles su imagen del primer cuadrante— con los cuadrantes hiperbólicos. En particular, dados dos cuadrantes hiperbólicos existe una única transformación hiperbólica que manda a uno en el otro.* □

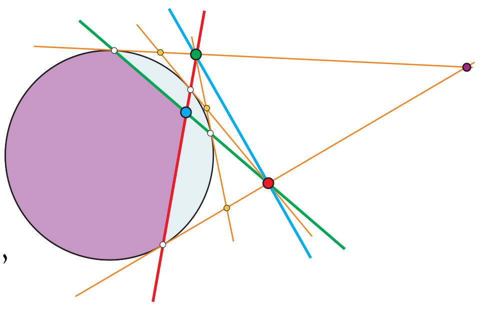
Formalmente, nos adelantamos al usar indistintamente los términos matriz y trans-

formación en las dos frases del teorema. Pero ya estamos en posición de concluir con soltura su equivalencia.

Proposicion 8.5 Hip $(2) = \mathcal{H}(2)$. *Es decir, cualquier transformación* $\mathsf{g} \in \mathrm{Pr}(2)$ *que preserva el círculo (tal que* $\mathsf{g}\left(\mathbb{S}^1\right) = \mathbb{S}^1$*) está determinada por una matriz* $\mathrm{B} \in \mathcal{H}(2)$.

Demostración. Ya vimos que $\mathcal{H}(2) \subset$ Hip (2) en (8.7), demostremos ahora la orta contención. Consideremos la terna ortogonal canónica $[\mathbf{e}_1]$, $[\mathbf{e}_2]$, $[\mathbf{e}_3]$; añadámosle las líneas hiperbólicas $[\mathbf{e}_1]^\perp$ y $[\mathbf{e}_2]^\perp$ así como las cuatro *tangentes* (*i.e.*, líneas polares) a sus puntos al infinito (hiperbólico, por supuesto). En esta configuración aparecen sólo cuatro nuevos puntos de las intersecciones de las tangentes (pues la definición de terna ortogonal obliga a que las seis líneas consideradas se partan en dos tercias concurrentes en los puntos espaciales de la terna). En nuestro caso, estos cuatro nuevos puntos son $(\pm 1, \pm 1)$, que podemos identificar naturalmente con los cuatro cuadrantes, de tal manera que al primero le corresponde el $(1, 1)$. Así que al primer cuadrante lo podemos identificar con la cuarteta de puntos en $\mathbb{P}^2$: $[\mathbf{e}_0]$, $[\mathbf{e}_1]$, $[\mathbf{e}_2]$, $[\mathbf{e}_3]$ (donde $\mathbf{e}_0 = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$), que fue nuestra cuarteta de batalla en las lides proyectivas y que llamábamos la cuarteta canónica.





Si $\mathsf{g} \in \mathrm{Pr}(2)$ preserva el círculo unitario, manda a sus tangentes en tangentes. Manda entonces las cuatro tangentes consideradas en cuatro equivalentes que dan información más que suficiente para demostrar que $\mathsf{g}([\mathbf{e}_1])$, $\mathsf{g}([\mathbf{e}_2])$, $\mathsf{g}([\mathbf{e}_3])$ es una terna ortogonal y que $\mathsf{g}([\mathbf{e}_0])$ identifica a uno de sus cuadrantes. Por el teorema anterior (la primera aseveración), existe $\mathrm{B} \in \mathcal{H}(2)$ que le hace lo mismo a la cuarteta canónica. Por el Teorema de 4 en 4, g y (la que define) B son la misma transformación. □

Tenemos también el subgrupo de transformaciones hiperbólicas que preservan orientación, que denotaremos $\mathcal{H}^+(2)$. Es fácil ver que es el de matrices con determinante positivo y más precisamente, que

$$\mathcal{H}^+(2) = \{\mathrm{B} \in \mathcal{H}(2) \mid \det \mathrm{B} = 1\}.$$

Para este grupo hiperbólico *orientado*, el teorema equivalente al 8.4 es que $\mathcal{H}^+(2)$ está en correspondencia con las parejas "punto en línea orientada" y que actúa simple y transitivamente en ellas.

EJERCICIO 8.18 Encuentra la matriz de la transformación hiperbólica orientada que deja invariante al eje x, con todo y su orientación, y "traslada" el $(0, 0)$ al $(1/2, 0)$.

EJERCICIO 8.19 Encuentra una matriz hiperbólica que manda al origen, $[\boldsymbol{e}_3]$, en el punto $(1/2, 1/2)$ y a la recta $[\boldsymbol{e}_2]^{\perp}$ en la recta $x = y$.

EJERCICIO 8.20 Observa que las rectas hiperbólicas $(\eta : 2x + y = 1)$ y $(\xi : x + y = 1)$ son ortogonales. Identifica las cuatro matrices que mandan al eje x y al eje y respectivamente en ellas.

8.3.4 Rotaciones y ángulos

Veamos ahora ejemplos concretos, empezando por las transformaciones hiperbólicas orientadas que fijan el origen.

Como se debió concluir en el Ejercicio 8.10, las extensiones del origen $[\boldsymbol{e}_3]$ a una terna ortogonal tienen como puntos espaciales dos ortogonales en la recta al infinito, es decir, de la forma $b = [s : t : 0]$ y $a = [-t : s : 0]$. Al levantarlos a unitarios (dividir entre la raíz de su L-norma, que en este caso coincide con la euclidiana) de tal manera que, juntos, preserven la orientación, obtenemos una matriz

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathcal{H}^+(2).$$

Representa a la *rotacion por un ángulo* θ *alrededor del origen* (pensando en $\mathbb{H}^2 \subset \mathbb{R}^2 \subset \mathbb{P}^2$, de la manera canónica). Éstas son todas las transformaciones hiperbólicas orientadas que dejan fijo el origen, y hemos demostrado la ecuación (8.1), que $O(2) \subset \mathcal{H}^+(2)$ son las transformaciones que fijan al origen.

Para obtener las rotaciones en cualquier otro punto $c \in \mathbb{H}^2$, usamos el viejo truco de conjugar. Existe alguna matriz $B \in \mathcal{H}^+(2)$ que manda a c en $[\boldsymbol{e}_3]$; entonces la *rotación de un ángulo* θ *alrededor de* c debe ser

$$R_{c,\theta} := B^{-1} R_\theta B,$$

pues manda a c en $[\boldsymbol{e}_3]$; rota ahí, y luego regresa a c a su lugar original. Ya podemos definir ángulos en el plano hiperbólico usando a las rotaciones.

Sean η y ξ dos rectas hiperbólicas orientadas que se intersectan en $c \in \mathbb{H}^2$. Definimos el *ángulo de* η *a* ξ como la única $\theta \in [0, 2\pi)$ tal que $R_{c,\theta}$ manda a η en ξ con las orientaciones correspondientes. Ésta no es una definición muy útil, pues para medir hay que encontrar alguna matriz de $\mathcal{H}^+(2)$ que mande a c en el origen y ahí calcular. Sin embargo, veremos en la siguiente sección que hay una manera mucho más fácil de encontrar el *ángulo*.

EJERCICIO 8.21 Demuestra que el ángulo es invariante bajo transformaciones hiperbólicas.

EJERCICIO 8.22 Demuestra que $f \in \mathcal{H}^+(2)$ tiene un punto fijo en $\mathbb{H}^2$ si y sólo si f es una rotación.

EJERCICIO 8.23 Sea $f \in \mathcal{H}^+(2))$ la rotación de un ángulo $\pi/2$ en el punto $\left(1/\sqrt{2}, 0\right)$. Sin usar matrices, demuestra que permuta cíclicamente a los puntos $(1,0)$, $\left(1/\sqrt{2}, 1/\sqrt{2}\right)$, $(-1,0)$ y $\left(1/\sqrt{2}, -1/\sqrt{2}\right)$. Recuerda que una transformación proyectiva está determinada por su valor en cuatro puntos, así que f ya lo está. Calculando dónde se intersectan los lados opuestos de este cuadrilátero, observa que se pueden encontrar los valores precisos de $f(0,1)$ y $f(0,-1)$. ¿Cuáles son?

EJERCICIO 8.24 Usando el Teorema de 4 en 4 (ver ejercicio anterior), demuestra que dado un punto $c \in \mathbb{H}^2$ y dos rectas orientadas que pasan por él, hay una única $f \in \mathcal{H}^+(2)$ que deja fijo a c y manda una recta en la otra. (Ojo, si las rectas no son ortogonales no hay una permutación cíclica como en el ejercicio anterior; pero puedes echar mano de que por un punto pasa una única ortogonal a una recta dada).

8.3.5 Traslaciones

Una *traslación* (a semejanza de las otras dos geometrías) deja fija una recta orientada y preserva orientación; queda entonces determinada por a dónde manda un punto en ella (pues es la información suficiente para determinar un elemento de $\mathcal{H}^+(2)$). Como dos puntos en el plano hiperbólico, p_0 y p_1 digamos, determinan una recta $\eta = \langle p_0, p_1 \rangle$, entonces definen una traslación.

...Dibujo

Veamos un caso particular: trasladar al origen $(0,0)$ el punto $c = (c, 0)$ con $-1 < c < 1$ (donde estamos abusando de la notación c como punto o como número). La recta $[\boldsymbol{e}_2]^\perp$ con su orientación se queda fija y, por tanto, $\boldsymbol{e}_2$ debe ser la segunda columna. El punto espacial que nos falta es

$$a = c^\perp \cap [\boldsymbol{e}_2]^\perp = \left(\frac{1}{c}, 0\right) = [1 : 0 : c] \,.$$

Debemos ahora levantar y normalizar nuestra terna y obtenemos

$$\begin{aligned} \boldsymbol{u} \; &: = \frac{1}{\sqrt{|(1,0,c)|_L^2}} (1,0,c) = \frac{1}{\sqrt{1-c^2}} (1,0,c) \\ \boldsymbol{v} \; &: = \boldsymbol{e}_2 \\ \boldsymbol{w} \; &: = \frac{1}{\sqrt{-|(c,0,1)|_L^2}} (c,0,1) = \frac{1}{\sqrt{1-c^2}} (c,0,1) \,. \end{aligned}$$

Debemos entonces definir la traslación hiperbólica de $(0,0)$ a $(c,0)$ como:

$$T_c := \frac{1}{\sqrt{1-c^2}} \begin{pmatrix} 1 & 0 & c \\ 0 & \sqrt{1-c^2} & 0 \\ c & 0 & 1 \end{pmatrix} ,$$

que está bien (no hay que cambiar signos de $\mathbf{u}$ y $\mathbf{v}$), pues su determinante es 1.

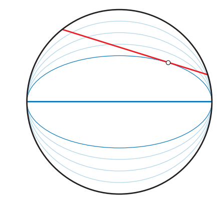
Todas las matrices T_c, con $-1 < c < 1$, forman un subgrupo de $\mathcal{H}^+(2)$ que llamaremos $\mathcal{T}_{\mathbf{e}_2}$ como acrónimo de *traslaciones en la recta* $[\mathbf{e}_2]^\perp$. Para ver que es un grupo basta observar que todos sus elementos tienen como puntos fijos a $(-1,0)$, $(1,0)$ y $[\mathbf{e}_2]$ (y por tanto preserva líneas verticales); en particular, la identidad es T_0. Podemos además orientar a $\mathcal{T}_{\mathbf{e}_2}$ en el sentido creciente de su parametro c. Si tomamos cualquier punto en la recta hiperbólica $[\mathbf{e}_2]^\perp$, es decir, de la forma $(s,0)$ con $-1 < s < 1$, sus imágenes bajo este grupo, *i.e.*, su *órbita,* es de nuevo la recta hiperbólica $[\mathbf{e}_2]^\perp$; pero podemos extender nuestro análisis de cómo actúa $\mathcal{T}_{\mathbf{e}_2}$ al resto de $\mathbb{P}^2$.

Si tomamos otro punto en $\mathbb{H}^2$ (el $(0,1/2)$ por ejemplo), su órbita es la mitad de una elipse con eje mayor el segmento del $(-1,0)$ al $(1,0)$. Para verlo, notemos que la ecuación de estas elipses es $x^2 + by^2 = 1$ para alguna $b > 1$ (en el ejemplo es $x^2 + 4y^2 = 1$); que, como elipse, se queda invariante bajo T_c, pues no es difícil ver que se tiene

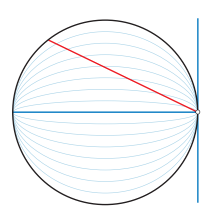
$$T_c^\top \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & -1 \end{pmatrix} T_c = \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

que, recordemos, es cómo cambia la matriz asociada al polinomio bajo la transformación proyectiva T_c. Así que si un punto cumple la ecuación, también lo hace su imagen bajo cualquier T_c. Notemos que lo anterior también funciona para $b \leq 1$, extendiéndose nuestro análisis a todo el plano proyectivo, dándonos elipses altas e hipérbolas que se parten en dos órbitas. Éstas también se orientan del $(-1,0)$ al $(1,0)$; excepto sus dos tangentes a $\mathbb{S}^1$, cuando $b = 0$, donde aparece $[\mathbf{e}_2]$ como punto singular. ¿Puedes hacer un dibujo de lo que pasa alrededor de $[\mathbf{e}_2]$?

Este grupo de traslaciones $\mathcal{T}_{\mathbf{e}_2} = \{T_c \mid -1 < c < 1\}$ nos servirá más adelante para medir distancias, pues las traslaciones deberían depender nada más de una recta orientada y una distancia. Pero por el momento podemos sacar dos conclusiones sobre distancias, aún sin haberla definido.



Las elipses interiores ($x^2 + by^2 = 1$ con $b > 1$) deben ser los conjuntos de puntos a una distancia fija de la recta $[\mathbf{e}_2]^\perp$, pues queremos que las transformaciones de $\mathcal{H}(2)$ preserven distancias. Así que una recta ultraparalela a $[\mathbf{e}_2]^\perp$ alcanza su distancia mínima a $[\mathbf{e}_2]^\perp$ en el punto en que es tangente a una de estas elipses y en ambas direcciones la distancia crece indefinidamente pues corta a todas las elipses exteriores.



Y como una recta paralela a $[\mathbf{e}_2]^\perp$ (que tiene punto al infinito $(1,0)$, digamos) cruza a todas las elipses, pues su tangente ahí es vertical, entonces tiene puntos tan cercanos a $[\mathbf{e}_2]^\perp$ como queramos. La distancia a $[\mathbf{e}_2]^\perp$ tiende a cero por un lado y a ∞ por el otro. Por supuesto, esto mismo pasa para cualquier par de rectas ultraparalelas, y paralelas, pues con el grupo podemos llevar a $[\mathbf{e}_2]^\perp$ en alguna de ellas.

EJERCICIO 8.25 Demuestra que para cualquier $B \in \mathcal{H}^+(2)$ existen $T_c \in \mathcal{T}_{\mathbf{e}_2}$ y rotaciones en el origen R_α, R_θ tales que $B = R_\theta T_c R_\alpha$. (Piensa geométricamente, y las coordenadas polares son útiles.)

EJERCICIO 8.26 Sean $a = (0,0)$, $b = (1/2, 0)$ y $c = (0, 1/2)$. Demuestra que

$$T_{a\leftarrow b} = \frac{1}{\sqrt{3}}\begin{pmatrix} 2 & 0 & -1 \\ 0 & \sqrt{3} & 0 \\ -1 & 0 & 2 \end{pmatrix} \quad \text{y} \quad T_{c\leftarrow a} = \frac{1}{\sqrt{3}}\begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix},$$

donde, en general, denotamos por $T_{q\leftarrow p}$ la traslación que manda a p en q (ambos en $\mathbb{H}^2$).

EJERCICIO 8.27 Con la notación del ejercicio anterior, demuestra que $T_{c\leftarrow a}T_{a\leftarrow b} \neq T_{c\leftarrow b}$ calculando qué le pasa al punto polar de la recta $\langle b, c\rangle$.

EJERCICIO 8.28 Siguiendo con la notación anterior, sea α el ángulo de la recta $\langle b, c\rangle$ (con dirección de b a c) con la recta $\langle b, a\rangle$ (hacia a). Demuestra que $\alpha < \pi/4$ (calculando que $\cos\alpha = 2/\sqrt{7}$) y por tanto que la suma de los ángulos internos del triángulo a, b, c suman menos que π.

EJERCICIO 8.29 Demuestra que $T_{c\leftarrow a}T_{b\leftarrow a} \neq T_{b\leftarrow a}T_{c\leftarrow a}$ calculando a dónde mandan al origen a; haz un dibujo.

EJERCICIO 8.30 Demuestra que

$$T_{c\leftarrow b} = \frac{1}{21}\begin{pmatrix} 24 & 3 & -12 \\ -11 & 24 & 16 \\ -16 & 12 & 29 \end{pmatrix}.$$

EJERCICIO 8.31 Demuestra que $T_{a\leftarrow c}T_{c\leftarrow b}T_{b\leftarrow a}$ es una rotación R_θ en a, donde $\theta = 2\alpha - \pi/2$. Concluye que las traslaciones hiperbólicas **no** son un grupo.

EJERCICIO 8.32 Vamos a cambiar de notación y a, b, c y d serán reales. Sea $p = (a, b) \in \mathbb{H}^2$; sean $c = \sqrt{a^2 + b^2}$ y $d = \sqrt{1 - c^2}$. ¿Cuál es la interpretación geométrica de c y d? Demuestra que la matriz de la traslación del origen $(0,0)$ a p es

$$\frac{1}{dc^2}\begin{pmatrix} a^2 + db^2 & ab(1-d) & ac^2 \\ ab(1-d) & da^2 + b^2 & bc^2 \\ ac^2 & bc^2 & c^2 \end{pmatrix}.$$

(Te conviene escribir las rotaciones en términos de a, b y c.)

EJERCICIO 8.33 Da un ejemplo de dos rotaciones cuya composición es una rotación. Piensa geométricamente.

EJERCICIO 8.34 Demuestra que $R_{q,\pi}R_{p,\pi} = T^2_{q\leftarrow p}$.

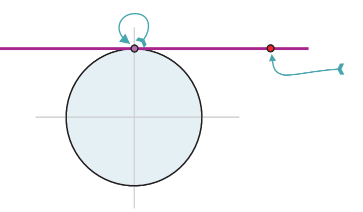
EJERCICIO 8.35 Demuestra que si $f \in \mathcal{H}^+(2)$ deja fijo un punto espacial η (*i.e.*, $f(\eta) = \eta$), entonces f es una traslación en $\eta^\perp$o una rotación de π en un punto de $\eta^\perp$.

8.3.6 Traslaciones horocíclicas

Las rotaciones dejan fijos puntos temporales y las traslaciones puntos espaciales. A las trasformaciones hiperbólicas orientadas que dejan fijo un único punto luz las llamaremos *traslaciones horocíclicas;* queremos ver que existen.

Las traslaciones dejan fijos dos puntos luz, pues al dejar fija una línea hiperbólica orientada dejan fijos sus puntos al infinito. Las traslaciones horocíclicas sólo dejan fijo un punto luz y al haz de paralelas por él lo "traslada", lo mueve en su dirección ortogonal sin dejar fijo ningun punto hiperbólico.

Para fijar ideas, consideremos al punto luz $p_0 = (0, 1)$ (o, proyectivamente, $p_0 = [0 : 1 : 1]$). Queremos construir una *traslación horocíclica* f *centrada en* p_0, es decir, que sólo fija p_0. Entonces debe mandar a su recta polar ($p_0^\perp$: $y = 1$) en sí misma, y por tanto debe mandar a $[e_1] \in p_0^\perp$ (que de hecho es su punto al infinito en la carta coordenada canónica) en un punto de la forma $x = (x, 1)$ para alguna $x \in \mathbb{R}$, $x \neq 0$ (abusaremos de la notación, usando igualdades para las inclusiones de las cartas coordenadas canónicas $\mathbb{R} \subset \mathbb{R}^2 \subset \mathbb{P}^2$, así que escribiremos $x = (x, 1) = [x : 1 : 1]$ y nos facilitará la escritura). Para fijar ideas, veamos el caso particular en que $[e_1]$ ($= \infty$) va al punto $2 = (2, 1)$; el caso general se deja como ejercicio.



La transformación f restringida a la recta proyectiva $p_0^\perp$ es una función de Moebius que deja fijo a p_0 ($= 0$). Si dejara fijo algún otro punto, éste sería espacial y tendríamos entonces una traslación hiperbólica común y corriente, pero queremos construir una traslación horocíclica. Entonces f restringida a $p_0^\perp$ es una traslación (de Moebius, con un solo punto fijo, recuérdese) que, junto con $f(\infty) = 2$, implica (véase Ejercicio 8.36) que

$$f(x) = \frac{2x}{x + 2}. \tag{8.8}$$

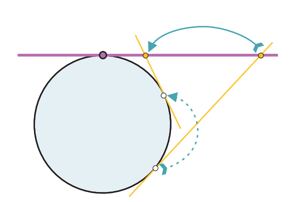
Ya tenemos un punto de la terna ortogonal que definirá a f. Sea

$$a = f([e_1]) = (2, 1).$$

Queremos ahora encontrar $b = f([e_2])$. Sabemos que b tiene que estar en la línea polar de a, que tiene como ecuación

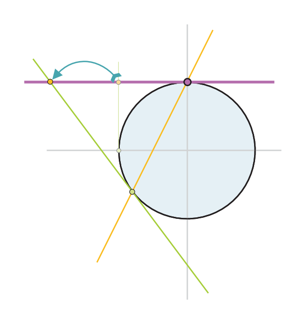
$$a^\perp : \ 2x + y = 1,$$

nos falta encontrar otra recta que pase por él.



La idea clave es que al saber cómo actúa f en la recta $p_0^\perp$, sabemos, en principio, cómo calcular la función f en el círculo luz. Dado un punto luz $p \in \mathbb{S}^1$, $p \neq p_0$, su tangente intersecta a la recta $p_0^\perp$ en el punto $p^\perp \cap p_0^\perp$ que corresponde a una x. Al aplicar f, tangentes deben ir en tangentes, así que $f(p)$ debe ser un pie de $f(x)$. Pero uno de los pies ya está fijo, es p_0, asi que $f(p)$ tiene que ser el otro. Los ejemplos importantes para nuestro problema son los puntos $p_1 := (1,0)$ y $p_{-1} := (-1,0)$ pues sus tangentes pasan por $[\mathbf{e}_2]$; son sus pies. El cálculo con p_{-1} sale más bonito. Hagámoslo.

$$\begin{aligned} p_{-1}^\perp \cap p_0^\perp &= (-1,1) = -1; \\ f(-1) &= -2 = (-2,1); \\ (-2,1)^\perp &: -2x + y = 1. \end{aligned}$$



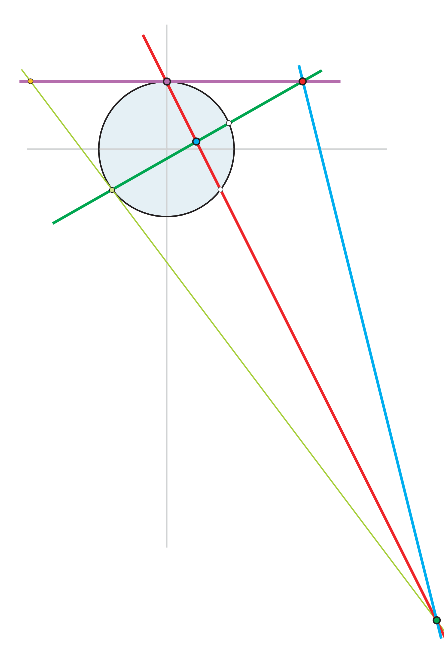
Y ahora debemos encontrar las intersecciones de esta recta con $\mathbb{S}^1$ para deducir el pie faltante de $(-2,1)$. Sustituyendo $y = 2x + 1$ en el polinomio de $\mathbb{S}^1$, se obtiene

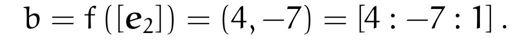
$$x^2 + (2x+1)^2 - 1 = x(5x+4).$$

Estas raíces son $x = 0$ (ya lo sabíamos) y $x = -4/5$, lo que nos da (al sustituir en la ecuación de $(-2,1)^\perp$ o de $\mathbb{S}^1$) $y = -3/5$, y por tanto que

$$f(p_{-1}) = f(-1,0) = -1/5\,(4,3) = [4:3:-5].$$

Y ahora sí, al intersectar su polar ($4x+3y = -5$) con $a^\perp$ ($2x+y = 1$), obtenemos a dónde debe ir $[\mathbf{e}_2]$:





$$b = f([\mathbf{e}_2]) = (4,-7) = [4:-7:1].$$

Finalmente, resolviendo otro sistema de 2×2, encontramos el punto temporal de la terna ortogonal como

$$c = a^\perp \cap b^\perp = 1/9\,(4,1) = [4:1:9].$$

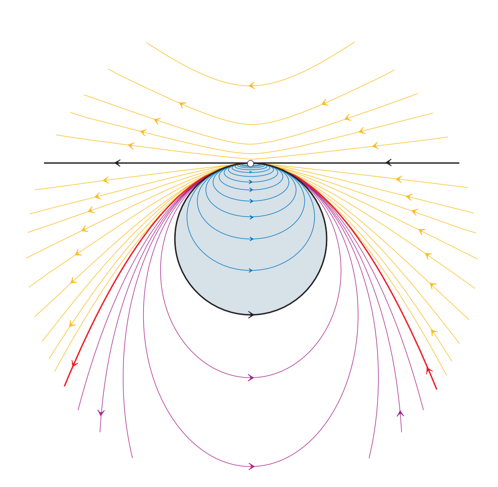
Levantamos a vectores L-unitarios, ajustamos los signos de los espaciales para eliminar la ambigüedad de 4, y obtenemos que la matriz de f tiene que ser

$$H_2 = \frac{1}{8}\begin{pmatrix} 8 & -4 & 4 \\ 4 & 7 & 1 \\ 4 & -1 & 9 \end{pmatrix}.$$

Es claro que el 2 que tomamos de ejemplo no tiene nada de particular y que para cualquier $a \neq 0$, se obtiene una traslación horocíclica que manda a $[\mathbf{e}_1]$ en $(a, 1)$. Dejamos como ejercicio ver que su matriz es

$$H_a = \frac{1}{2a^2}\begin{pmatrix} 2a^2 & -2a & 2a \\ 2a & 2a^2-1 & 1 \\ 2a & -1 & 2a^2+1 \end{pmatrix}. \tag{8.9}$$

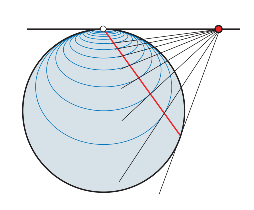
Cuando a tiende a ∞, H_a tiende a la identidad, I. Así que si la incluimos, obtenemos un subgrupo de $\mathcal{H}^+(2)$: las *traslaciones horocíclicas centradas en* $p_0 = [0:1:1]$, que denotaremos $\mathcal{T}_{p_0}$ (la notación extiende a puntos luz la de las traslaciones centradas en puntos espaciales) y que es isomorfo a las traslaciones de la recta real (en nuestro caso, identificada con $\mathbb{P}^1 - \{0\}$). Orientado de ∞ a 1, veremos a continuación cuáles son sus órbitas. A las que quedan dentro de $\mathbb{H}^2$ se les llama *horociclos*, y demostraremos que son las curvas que intersectan ortogonalmente a todas las rectas del haz paralelo centrado en p_0.



Las órbitas de T_{p_0} *son las cónicas de la familia que tiene como "círculo de curvatura" en el* $(0, 1)$ *al círculo unitario* $\mathbb{S}^1$. Además de ser tangentes a $\mathbb{S}^1$ en el $(0, 1)$: éste es el círculo que más se les parece infinitesimalmente ahí (ver Ejercicios 8.41 y 8.42). Para demostrarlo, las definimos por la familia de polinomios

$$P_b(x, y) = b\,x^2 + (y + b - 1)^2 - b^2,$$

donde $(0, 1 - b)$ es su centro. Esta familia tiene a la recta doble $(y - 1)^2$ para $b = 0$; consta de elipses para $b > 0$; incluye a la parábola $x^2 + 2y - 2$ cuando $b \to \infty$, y da hipérbolas para $b < 0$. Con un cálculo directo, aunque engorroso, se puede ver que son invariantes bajo las transformaciones H_a; es decir, que cumplen $\tilde{P}_b \circ H_a = \tilde{P}_b$, donde $\tilde{P}_b$ es su polinomio homogeneizado. Entonces, si un punto está en una de estas cónicas, toda su órbita también lo está.



Para $0 < b < 1$, la elipse $P_b(x, y)$ es un horociclo. Las tangentes a los horociclos en sus intersecciones con una recta hiperbólica que pasa por p_0 concurren en un punto de $p_0^\perp$, pues son la imagen bajo alguna H_a de las correspondientes horizontales que concurren en $[\mathbf{e}_1]$. A saber, tal recta es $[a]^\perp$ para algún $[a] \in p_0^\perp$.

Hemos exhibido a las traslaciones horocíclicas correspondiendo a las transformaciones de Moebius de una recta tangente al círculo luz que tienen como único punto fijo al de tangencia (ver Ejercicios 8.36 y 8.38). Veremos en la siguiente sección que esto no es casualidad pues nuestro razonamiento se extiende a cualquier transformación de Moebius que fije el punto de tangencia. Pero, por el momento, concluyamos con que ya hemos descrito todas las transformaciones hiperbólicas orientadas.

Teorema 8.6 *Sea* $f \neq \text{id}$ *una transfomación hiperbólica que preserva orientación; entonces* f *es una rotación, una traslación o una traslación horocíclica.*

Demostración. Nótese primero que hay que evitar la identidad pues ésta califica para las tres clases. Sea $B \in \mathcal{H}^+(2)$ la matriz asociada a f. Puesto que toda matriz de 3×3 tiene al menos un vector propio, entonces $f : \mathbb{P}^2 \to \mathbb{P}^2$ tiene un punto fijo. Si este punto fijo es temporal entonces es una rotación.

Supongamos ahora que el punto fijo, a digamos, es espacial. Sean q_1 y q_2 los pies de a. Como $f(a) = a$, estos pies van en sí mismos. Si se quedan fijos, *i.e.*, $f(q_i) = q_i$, entonces f es una traslación. La única otra posibilidad es que f los permute. En este caso, consideremos la función f restringida a la recta $a^\perp$. Es una transformación de Moebius ahí, que permuta dos puntos pero manda a los dos segmentos (el temporal y el espacial) en sí mismos. Por el Ejercicio 8.43, tiene un punto temporal fijo y entonces es una rotación de π en ese punto.

Finalmente, si el único punto fijo de f es luz, entonces f es por definición una traslación horocíclica. □

EJERCICIO 8.36 Demuestra que si una transformacion de Moebius tiene como único punto fijo al 0 entonces es de la forma $\frac{ax}{x+a}$ para algún $a \neq 0$, que es la imagen del infinito.

EJERCICIO 8.37 Como calentamiento para el caso general, encuentra la matriz de la traslación horocíclica alrededor de $p_0 = (0, 1)$ que manda a $[\mathbf{e}_1]$ en el punto $(1, 1)$. Sigue los pasos del ejemplo en el texto.

EJERCICIO 8.38 Demuestra que la matriz de (8.9) es la traslación horocíclica centrada en $p_0 = (0, 1)$ que manda a $[\mathbf{e}_1]$ en el punto $(a, 1)$. (Hay una demostración constructiva y, al menos, una teórica donde necesitas el siguiente ejercicio; suponlo cierto.)

***EJERCICIO 8.39** Demuestra que el polinomio característico de H_a es $-(\lambda - 1)^3$.

EJERCICIO 8.40 Usando al Ejercicio 8.38, encuentra una formúla para expresar cualquier traslación horocíclica.

EJERCICIO 8.41 Considera a la elipse canónica $b^2x^2 + a^2y^2 = a^2b^2$ parametrizada por $\theta \mapsto p(\theta) = (a\cos\theta, b\,\text{sen}\,\theta)$. Para cada $\theta \neq 0$, sea C_θ el círculo que pasa por los puntos $p(\theta), p(-\theta)$ y $p(0) = (a, 0)$. Demuestra que la abcsisa del centro del círculo C_θ tiende a $\left(a^2 - b^2\right)/a$ cuando θ tiende a 0. El círculo límite es el *círculo de curvatura* de la elipse en el punto $(a, 0)$. Con la misma idea se puede definir el círculo de curvatura en cualquier otro punto y para otras cónicas.

EJERCICIO 8.42 Demuestra que $\mathbb{S}^1$ es el círculo de curvatura de la elípse $P_b(x, y) = b\,x^2 + y^2 + 2(b-1)y - 2b + 1$, con $b > 0$, en el punto $(0, 1)$.

EJERCICIO 8.43 Demuestra que una transformación de Moebius que manda a un segmento en sí mismo, pero intercambia sus extremos, tiene un punto fijo.

8.3.7 Transformaciones hiperbólicas y de Moebius (su isomorfismo)

Las transformaciones de Moebius han aparecido como herramienta en nuestro estudio de las transformaciones hiperbólicas, y como ya dijimos, no es por casualidad. Esta sección estará dedicada a demostrar su estrechísima relación.

Teorema 8.7 *El grupo de transformaciones hiperbólicas es naturalmente isomorfo, como grupo, al de transformaciones de Moebius, es decir,*

$$\mathcal{H}(2) \simeq \Pr(1).$$

La idea básica es identificar el círculo al infinito de $\mathbb{H}^2$, con la recta proyectiva $\mathbb{P}^1$. De hecho, ya lo hicimos con bastante éxito en la sección anterior tomando el punto fijo $p_0 = (0, 1)$, y definiendo geométricamente una biyección $\mathbb{S}^1 \longleftrightarrow \mathbb{P}^1$ al intersectar $p_0^\perp$ con las tangentes, que podemos escribir:

$$\begin{aligned} \varphi \quad &: \quad \mathbb{S}^1 \to \mathbb{P}^1 \left(= p_0^\perp\right) \\ \varphi(p) \quad &= \quad \begin{cases} p^\perp \cap p_0^\perp & \text{para} \quad p \neq p_0 \\ p_0 & \text{para} \quad p = p_0 \end{cases} . \end{aligned} \tag{8.10}$$

Nos conviene encontrar su expresión explícita en coordenadas pues la usaremos hacia el final de la demostración.

Estamos identificando a $\mathbb{P}^1$ con la recta $y = 1$ en $\mathbb{R}^2$, que en $\mathbb{P}^2$ está dada por la ecuación homogénea $y = z$; y la manera natural de hacerlo es mediante la inclusión que en coordenadas homogéneas toma la forma $[\lambda : \mu] \mapsto [\lambda : \mu : \mu]$. Dado $p = (\alpha, \beta) \in \mathbb{S}^1$, queremos encontrar el punto $p^\perp \cap p_0^\perp$. Sabemos que $p^\perp$ está determinado por la ecuación lineal homogénea $\alpha x + \beta y - z = 0$. Al sustituir $z = y$, obtenemos la ecuación

$$\alpha x + (\beta - 1) y = 0,$$

cuya solución es $[x : y] = [1 - \beta : \alpha]$; es decir $p^\perp \cap p_0^\perp = [1 - \beta : \alpha : \alpha]$. Esto no tiene sentido para $\beta = 1$, pues $(\alpha, \beta) \in \mathbb{S}^1$ implica que $\alpha = 0$ y no se debe, ni se puede, hablar de $[0 : 0]$. Sin embargo, si suponemos que $\beta \neq \pm 1$ se tiene, puesto que $\alpha^2 + \beta^2 = 1$, que

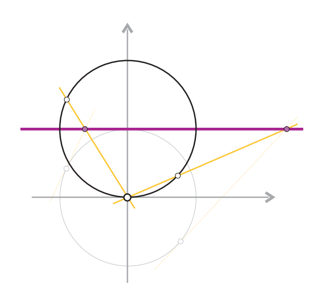
$$\begin{aligned} [1 - \beta : \alpha] &= [(1 + \beta)(1 - \beta) : (1 + \beta)\alpha] \\ &= \left[1 - \beta^2 : (1 + \beta)\alpha\right] = \left[\alpha^2 : (1 + \beta)\alpha\right] = [\alpha : 1 + \beta]. \end{aligned}$$

Tenemos entonces que nuestra identificación del círculo unitario con la línea proyectiva tiene que ser

$$\varphi : \mathbb{S}^1 \to \mathbb{P}^1$$
$$\varphi(\alpha, \beta) = \begin{cases} [1 - \beta : \alpha] & \text{para} \quad \beta \neq 1 \\ [\alpha : 1 + \beta] & \text{para} \quad \beta \neq -1 \end{cases}, \tag{8.11}$$

que aunque esté expresada de manera distinta en dos grandes pedazos (cada uno el círculo menos un punto), se tiene que en su intersección coinciden.

Un corolario importante de esta "talacha" que acabamos de hacer es que la identificación de $\mathbb{S}^1$ con $\mathbb{P}^1$ que estamos usando coincide con la clásica **proyección estereográfica**. Pues nótese que la última fórmula se puede pensar en dos tiempos. En su versión en $\mathbb{R}^2$, la función $(\alpha, \beta) \mapsto (\alpha, \beta + 1)$ es simplemente trasladar una unidad hacia arriba. Y luego se proyectiviza a cada punto del círculo trasladado, que se puede interpretar como tomar la intersección de su recta por el origen (en donde quedó el polo sur) con la recta canónica $y = 1$ identificada con $\mathbb{P}^1$ (en el origen hay que cambiar de definición y tomar el límite de las líneas: la tangente que intersecta en el punto al infinito). Así que la función φ coincide con la proyección desde el polo sur a la recta horizontal por el centro del círculo, justo la llamada proyección estereográfica.



Una vez establecida la correspondencia entre $\mathbb{S}^1$ y $\mathbb{P}^1$, conviene simplificar nuestra notación. Volvamos a pensar en $\mathbb{P}^1$ como la recta real extendida y para $x \in \mathbb{P}^1 = \mathbb{R} \cup \{\infty\}$ (que también pensaremos como $x \in p_0^\perp$) denotemos:

$$p_x := \varphi^{-1}(x) \in \mathbb{S}^1,$$

en concordancia con el uso que dimos a p_0, p_1 y p_∞ en la sección anterior.

Tenemos que demostrar ahora que al identificar así a $\mathbb{S}^1$ con $\mathbb{P}^1$, las transformaciones hiperbólicas y las transformaciones de Moebius se corresponden. Es decir, que para cualquier $f \in \mathcal{H}(2)$, la transformación $\mu_f : \mathbb{P}^1 \to \mathbb{P}^1$ definida por la ecuación

$$\mu_f(x) = \varphi(f(p_x))$$

es de Moebius. Y al revés, que es como vamos a proceder, que para cualquier transformación de Moebius μ, existe una transformación hiperbólica f_μ tal que

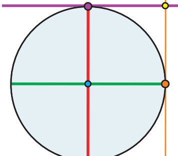
$$f_\mu(p_x) = p_{\mu(x)}.$$

Ya hemos avanzado en esto. Para las transformaciones de Moebius que fijan únicamente al 0, ya lo hicimos. Y las mismas ideas funcionan para el caso general. Demostremos que dada una transformación de Moebius μ, entonces la función

$$f_\mu : \mathbb{S}^1 \to \mathbb{S}^1$$
$$f_\mu(p_x) = p_{\mu(x)}$$

se extiende a una transformación hiperbólica. De nuevo, el punto será encontrar una terna ortogonal adecuada.

Sabemos que μ está determinada por sus valores en $0, 1$ e ∞. Sean entonces $q_0 := p_{\mu(0)}$, $q_1 := p_{\mu(1)}$ y $q_\infty := p_{\mu(\infty)}$ los puntos correspondientes en $\mathbb{S}^1$. Como la terna ortogonal canónica se define en términos de la "terna luz" canónica, p_0, p_1, p_∞, como sigue



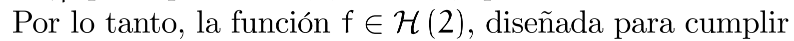
$$\begin{aligned} [\mathbf{e}_1] &= p_0^\perp \cap p_\infty^\perp; \\ [\mathbf{e}_2] &= [\mathbf{e}_1]^\perp \cap p_1^\perp; \\ [\mathbf{e}_3] &= [\mathbf{e}_2]^\perp \cap [\mathbf{e}_1]^\perp; \end{aligned}$$

Entonces definimos

$$\begin{aligned} a &:= q_0^\perp \cap q_\infty^\perp; \\ b &:= a^\perp \cap q_1^\perp; \\ c &:= b^\perp \cap a^\perp. \end{aligned}$$

Esta terna se levanta a cuatro ternas L-ortogonales; pero es fácil ver que sólo una de ellas manda a nuestros tres puntos como queremos. Observemos que el cuarto punto de la cuarteta proyectiva canónica $[\boldsymbol{e}_0] = [\boldsymbol{e}_1 + \boldsymbol{e}_2 + \boldsymbol{e}_3]$ también se define como $[\boldsymbol{e}_0] = p_0^\perp \cap p_1^\perp$, así que hay que mandarlo a $q_0^\perp \cap q_1^\perp$ para que $f \in \mathcal{H}(2) \subset \Pr(2)$ quede determinada.

Por lo tanto, la función $f \in \mathcal{H}(2)$, diseñada para cumplir



$$f(p_x) = p_{\mu(x)} \; (= q_x) \quad \text{para} \quad x = 0, 1, \infty, \tag{8.12}$$

existe y es única. Pero nos falta demostrar que su extensión a todo $\mathbb{S}^1$ es como la transformación de Moebius μ manda; es decir, que se cumple la ecuación anterior para cualquier $x \in \mathbb{P}^1$. Esto, en notación funcional, equivale a probar que $f \circ \varphi^{-1} = \varphi^{-1} \circ \mu$, pues recuérdese que $\varphi^{-1}(x) = p_x$; así que nos queda por demostrar que

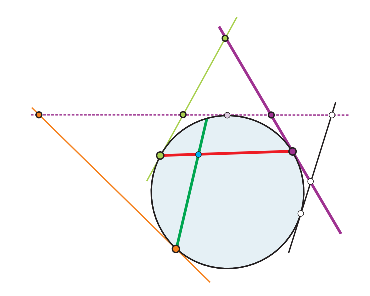
$$\varphi \circ f \circ \varphi^{-1} = \mu. \tag{8.13}$$

Y esto es equivalente a probar que $\varphi \circ f \circ \varphi^{-1}$ **es una transformación proyectiva** de la recta $p_0^\perp$ en sí misma, pues ya coincide con μ en la terna canónica (8.12).

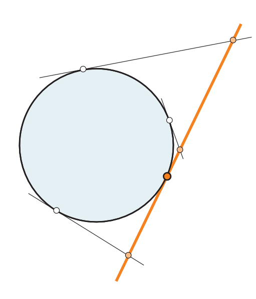
El caso en que $f(p_0) = p_0$, es decir, que $\mu(0) = 0$, es fácil, y casi se vio en la sección anterior. Pues entonces f es una función proyectiva de la recta $p_0^\perp$ en sí misma que, es lo que bastará demostrar, es igual a $\varphi \circ f \circ \varphi^{-1}$, porque f preserva tangencia (o, en general, polaridad). Veamos. Para cualquier $x \neq 0$ se tiene que $f \circ \varphi^{-1}(x) = f(p_x) \neq f(p_0) = p_0$ y entonces, usando la definición original de φ (8.10), se tiene que

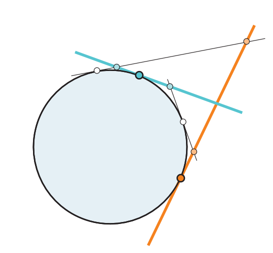
$$\begin{aligned} \varphi(f(p_x)) &= f(p_x)^\perp \cap p_0^\perp = f(p_x)^\perp \cap f(p_0)^\perp \\ &= f\left(p_x^\perp\right) \cap f\left(p_0^\perp\right) = f\left(p_x^\perp \cap p_0^\perp\right) = f(x). \end{aligned} \tag{8.14}$$

El caso general en que $f(p_0) \neq p_0$ requiere más trabajo, pues f manda, proyectivamente, a la línea $p_0^\perp$ en otra línea (a saber, $q_0^\perp = f(p_0)^\perp = f\left(p_0^\perp\right)$), y para obtener $\varphi \circ f \circ \varphi^{-1}$ hay que "bajarla" o "untarla" al círculo y volverla a "trepar" a $p_0^\perp$. Para $x \neq 0$, lo que sí sabemos es que tanto $f(x)$ $\left(\in q_0^\perp\right)$ como $\varphi(f(p_x))$ $\left(\in p_0^\perp\right)$ están en la línea $f(p_x)^\perp$; lo primero porque $x \in p_x^\perp$ y lo segundo por definición de φ. Esta situación cae en el esquema general de una correpondencia natural que se establece entre las líneas tangentes al círculo, que llamaremos *correspondencia tangencial*. Nos conviene liberarnos un rato del problema que traemos, con todo y su notación que nos obstruye la visión, y establecer esto último con calma y sin la presión de demostrar.



Notemos que lo único que tiene p_0 de especial es que facilita los cálculos, pero para cualquier otro punto en $\mathbb{S}^1$, q digamos, hay una biyección natural $\varphi_q : \mathbb{S}^1 \to q^\perp$ que deja fijo a q y manda cualquier otro $p \in \mathbb{S}^1$ a $p^\perp \cap q^\perp$ (compárese con (8.10)). Su inversa es simplemente mandar un punto $y \in q^\perp$ a **su otro pie** en $\mathbb{S}^1$ (q es el pie fijo, y se queda ahí, parado); es lo que habíamos llamado "untar" a $q^\perp$ en el círculo. Lo podemos volver a levantar a cualquier otra línea tangente —$p^\perp$ digamos, con $p \in \mathbb{S}^1$—, y obtenemos una biyección entre las dos líneas tangentes: su *correspondencia tangencial*, $\varphi_p \circ \varphi_q^{-1}$. Dicho de otra manera y genéricamente, $y \in q^\perp$ corresponde tangencialmente a $x \in p^\perp$ si y sólo si la línea $\langle x, y\rangle$ es tangente a $\mathbb{S}^1$. Nótese que es importante pensar el punto en su línea, pues si $z = p^\perp \cap q^\perp$ (que es el único caso para que lo genérico no funcione) entonces $z \in p^\perp$ corresponde a $q \in q^\perp$ mientras que $z \in q^\perp$ corresponde a $p \in p^\perp$ (son sus dos pies).

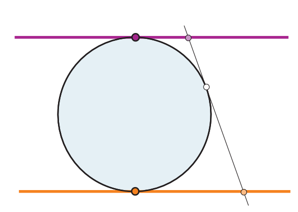




Puesto que la correspondencia tangencial surge, en el fondo, de identificaciones con un objeto único —el círculo—, es fácil ver que conmutan, es decir, que la composición de dos de ellas es otra; y es en ese sentido que usamos el término en singular. El punto medular para concluir con el teorema es demostrar que estas correspondencias entre líneas proyectivas son muy bien portadas.

Proposicion 8.8 *La correspondencia tangencial entre dos tangentes a $\mathbb{S}^1$ es una transformación proyectiva.*

Demostración. Es claro que las transformaciones hiperbólicas preservan la correspondencia tangencial, pues esta no sólo está definida entre tangentes sino en base a tangentes (de $\mathbb{S}^1$, por supuesto). Por tanto, podemos escoger nuestro par de rectas y, por conjugación —o contagio grupal—, si probamos la proposición ahí, será cierta para cualquier par.



Pensando en $\mathbb{R}^2$, consideremos a $p_0^\perp$ y $p_\infty^\perp$ parametrizadas, ambas, por la primera coordenada (la segunda es 1 o -1 respectivamente). Dado $p = (\alpha, \beta) \in \mathbb{S}^1$, vimos al principio de la sección que le correponde $(1-\beta)/\alpha$ en $p_0^\perp$, pero además vimos (8.11) que para $\beta \neq \pm 1$ esto es igual a $\alpha/(1+\beta)$. Sustituyendo $y = -1$ en la ecuación de $p^\perp$ ($\alpha x + \beta y = 1$) se ve que a p le corresponde $(1+\beta)/\alpha$ en $p_\infty^\perp$. De tal manera que la correspondencia tangencial entre $p_0^\perp$ y $p_\infty^\perp$ se expresa en parámetros como $x \leftrightarrow 1/x$, que es de Moebius. $\square$

Una observación que no podemos dejar pasar es que esto último, junto con el comentario que sigue a (8.11), implica que la proyección estereográfica del eje x a $\mathbb{S}^1$ desde el polo sur, seguida de la proyección estereográfica desde el polo norte de regreso a $\mathbb{R}$, es la función $x \mapsto 1/x$.

Ahora sí, tenemos todos los elementos para concluir la demostración del teorema.

Demostración. [$\mathcal{H}(2) \simeq \Pr(1)$] Si regresáramos a la frase donde nos desviamos de la demostración,(despuesito de (8.13)) concluiríamos que $\varphi \circ f \circ \varphi^{-1}$ es f seguida de la correspondencia tangencial de $f\left(p_0^\perp\right)$ a $p_0^\perp$, y que por tanto $\varphi \circ f \circ \varphi^{-1}$ es proyectiva; pero aprovechemos la distracción para recapitular y concluir de un solo golpe.

Para una transformación proyectiva μ de $p_0^\perp$ en sí misma construimos, usando la identificación $\varphi : \mathbb{S}^1 \to p_0^\perp$, una transformación hiperbólica $f \in \mathcal{H}(2)$ tal que $\varphi \circ f \circ \varphi^{-1}$ coincide con μ en tres puntos. Estábamos en demostrar que $\varphi \circ f \circ \varphi^{-1}$ también es proyectiva pues, por el Teorema de 3 en 3, esto es equivalente a que coincida con μ. Lo haremos ahora en general, mimetizando el caso que ya habíamos concluido con éxito (8.14). ...Dibujo

Sea $q = f^{-1}(p_0) \in \mathbb{S}^1$ para que $f(q) = p_0$ y f sea una transformación proyectiva de $q^\perp$ en $p_0^\perp$. Entonces, para cualquier $x \in p_0^\perp$ tal que $q \neq p_x = \varphi^{-1}(x) \in \mathbb{S}^1$, se tiene, puesto que $f(p_x) \neq p_0$, que

$$\begin{aligned} \varphi \circ f \circ \varphi^{-1}(x) &= \varphi(f(p_x)) = f(p_x)^\perp \cap p_0^\perp \\ &= f(p_x)^\perp \cap f(q)^\perp = f\left(p_x^\perp\right) \cap f\left(q^\perp\right) = f\left(p_x^\perp \cap q^\perp\right). \end{aligned}$$

Nótese que $p_x^\perp \cap q^\perp$ es la imagen de $x \in p_0^\perp$ bajo la correspondencia tangencial de $p_0^\perp$ a $q^\perp$; entonces hemos expresado $\varphi \circ f \circ \varphi^{-1}$ como la composición de dos transformaciones proyectivas y por tanto es proyectiva. Concluimos entonces que

$$\varphi \circ f \circ \varphi^{-1} = \mu. \tag{8.15}$$

De aquí en adelante nada más falta comentar detalles. Obsérvese primero que en el último párrafo no se usó sino hasta la última línea de dónde viene f; es cierto entonces que para cualquier $f \in \mathcal{H}(2)$ la transformación $\varphi \circ f \circ \varphi^{-1}$ que va de $p_0^\perp$ en sí misma, definida por conjugación con φ, es proyectiva. Así que, identificando a $p_0^\perp$ con $\mathbb{P}^1$, como lo hicimos con cuidado al principio de la sección, tenemos una función

$$\begin{aligned} \mathcal{H}(2) &\to \Pr(1) \\ f &\mapsto \varphi \circ f \circ \varphi^{-1} \end{aligned}$$

cuya inversa es la que se llevó la parte constructiva de la demostración, y que (se deduce fácilmente de (8.15) despejando f) se escribe

$$\begin{aligned} \Pr(1) &\to \mathcal{H}(2) \\ \mu &\mapsto \varphi^{-1} \circ \mu \circ \varphi. \end{aligned}$$

Tenemos entonces la biyección natural del teorema. Pero además es un *isomorfismo de grupos* pues en ambos sentidos preserva la composición (justo por estar definidas como conjugación). Por ejemplo,

$$g \circ f \mapsto \varphi \circ (g \circ f) \circ \varphi^{-1} = \left(\varphi \circ g \circ \varphi^{-1}\right) \circ \left(\varphi \circ f \circ \varphi^{-1}\right).$$

□

EJERCICIO 8.44 Encuentra las matrices de las transformaciones hiperbólicas que mandan a la terna luz canónica p_0, p_1, p_∞ respectivamente en

$$\begin{array}{ccc} \left[1 : \sqrt{3} : 2\right], & \left[1 : -\sqrt{3} : 2\right], & [-1 : 0 : 1]; \\ (0,1), & (-1,0), & (1,0); \\ [4 : 3 : 5], & [-4 : 3 : 5], & [0 : -1 : 1]. \end{array}$$

EJERCICIO 8.45 Demuestra que la función inversa de φ (8.11) tiene la forma

$$\varphi^{-1}([\lambda : \mu]) = \frac{1}{\lambda^2 + \mu^2}\left(2\lambda\mu, \mu^2 - \lambda^2\right).$$

EJERCICIO 8.46 Recuerda de la Sección 6.2.5 la clasificación de transformaciones de Moebius; ¿a qué tipo de transformaciones hiperbólicas corresponden?

EJERCICIO 8.47 Demuestra que si $f \in \mathcal{H}(2)$ invierte orientación, entonces es una reflexión (deja fija, punto a punto, a una línea hiperbólica llamada su espejo, e invierte sus dos lados) o es un paso (una reflexión seguida de una traslación en el espejo).

EJERCICIO 8.48 Demuestra que dos líneas hiperbólicas son ortogonales si y sólo si sus extremos forman una cuarteta armónica (identificando al círculo al infinito con $\mathbb{P}^1$).

EJERCICIO 8.49 ¿Qué tipo de transformación se obtiene de la composición de dos reflexiones?

8.4 Métrica

Vamos ya a acabar de definir las nociones básicas de ángulo y distancia en $\mathbb{H}^2$ de tal manera que el grupo las preserve. Conviene pensar y hacer las cosas formales en el modelo de la cazuela hiperbólica $\mathbb{H}^2 \subset \mathbb{R}^3$.

8.4.1 Ángulos

Ya definimos ángulo con base en las rotaciones, que debía ser llamado el *ángulo orientado*, pues se definió como el ángulo de la rotación que lleva una recta en otra y toma valores entre $-\pi$ y π o bien entre 0 y 2π. No es simétrico sino antisimétrico, pues la inversa de esa rotación (menos el ángulo) nos regresa a la original. Y esta noción que en el origen hiperbólico, $\mathbf{e}_3$, coincide con la euclidiana se difundió a todo $\mathbb{H}^2$ por medio del grupo $\mathcal{H}(2)$. Veremos ahora cómo calcular fácilmente su versión simétrica, el ángulo con valores entre 0 y π, para dos rectas orientadas.

En la Sección 8.3 vimos la correspondencia natural entre vectores L-unitarios espaciales y rectas hiperbólicas orientadas; denotémosla con $\mathbf{u} \leftrightarrow \mathbf{u}^\perp$, donde $\mathbf{u}^\perp$ denota

ahora una recta hiperbólica orientada. En el "origen hiperbólico", $\boldsymbol{e}_3$, esta correspondencia es la euclidiana; una recta orientada por $\boldsymbol{e}_3$ es (la intersección con) un plano vertical y su vector asociado es su ortogonal (con tercera coordenada igual a 0). Y como para estos vectores el producto interior euclidiano corresponde al de Lorentz-Minkowski se tiene que

$$\measuredangle(\mathbf{u}^{\perp}, \boldsymbol{v}^{\perp}) = \arccos(\mathbf{u} \bullet_{\mathsf{L}} \boldsymbol{v}) \tag{8.16}$$

correponde a nuestra definición previa usando al grupo $\mathrm{SO}(2) \subset \mathcal{H}(2)$.

Simplemente hay que extender esta fórmula a cualquier par de rectas hiperbólicas orientadas $\mathbf{u}^{\perp}$ y $\boldsymbol{v}^{\perp}$ concurrentes para definir su ángulo. Expresa lo que queremos geométricamente, pues al actuar por elementos de $\mathcal{H}(2)$ se queda invariante, así que corresponde a llevar su intersección al origen y ahí medir con el grupo. En particular, si dos rectas hiperbólicas $\mathbf{u}^{\perp}$ y $\boldsymbol{v}^{\perp}$ se intersectan se tiene que $|\mathbf{u} \bullet_{\mathsf{L}} \boldsymbol{v}| \leq 1$ y $\measuredangle(\mathbf{u}^{\perp}, \boldsymbol{v}^{\perp}) \in [0, \pi]$, donde los extremos se dan si y sólo si $\mathbf{u} = \pm\boldsymbol{v}$.

Conviene extender la fórmula para calcular ángulos a vectores no necesariamente unitarios. Para cualquier $\boldsymbol{v} \in \mathbb{R}^3$, denotemos por

$$\|\boldsymbol{v}\|_{\mathsf{L}} := \sqrt{\left|\,|\boldsymbol{v}|^2_{\mathsf{L}}\right|}$$

a su L-*tamaño*. Nótese que el término "L-norma" debería aplicarse aquí en estricta analogía con el caso euclidiano, y deberíamos haber usado "L-norma cuadrada" para $|\boldsymbol{v}|^2_{\mathsf{L}}$; pero como esta última es la más usada preferimos simplificar la terminología y usar y acuñar, hasta ahora que lo necesitamos, un nuevo término para la raiz del valor absoluto de la L-norma, que sólo sirve para *normalizar*, esto es, para hacer L-unitario a un vector no luz.

Se tiene entonces que si $\mathbf{u}$ y $\boldsymbol{v}$ son vectores espaciales cuyas rectas polares son (hiperbólicamente) concurrentes,

$$\measuredangle(\mathbf{u}^{\perp}, \boldsymbol{v}^{\perp}) = \arccos\left(\frac{\mathbf{u} \bullet_{\mathsf{L}} \boldsymbol{v}}{\|\mathbf{u}\|_{\mathsf{L}}\,\|\boldsymbol{v}\|_{\mathsf{L}}}\right). \tag{8.17}$$

Veamos un ejemplo sencillo pero iluminador. Pensando en el modelo proyectivo ($\mathbb{H}^2 \subset \mathbb{R}^2 \subset \mathbb{P}^2$), consideremos la familia de rectas con una pendiente euclidiana dada y calculemos su ángulo hiperbólico respecto al eje x, $\boldsymbol{e}_2^{\perp}$. Sea (a, b), con $a > 0$ (para simplificar expresiones, pero $a \neq 0$ funciona), un vector euclidianamente unitario (es decir, tal que $a^2 + b^2 = 1$), de tal manera que el haz (euclidianamente) ortogonal a (a, b) está dado por las ecuaciones

$$ax + by = ac,$$

donde c, variable, parametriza su intersección con $\boldsymbol{e}_2^{\perp}$, que es precisamente $(c, 0)$. Sea $\mathbf{u}_c = (a, b, ac)$, así que estas rectas son sus (L-) polares y son hiperbólicas para $|c| < 1/a$. Se tiene que para estos valores de c:

$$\frac{\mathbf{e}_2 \bullet_L \mathbf{u}_c}{\|\mathbf{u}_c\|_L} = \frac{b}{\sqrt{1-a^2c^2}}.$$

Dibujo: grafica de $\rho(c)$

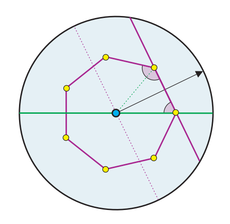
Llamemos $\rho(c)$ a esta función con dominio $-1/a < c < 1/a$. Notemos primero que si $b = 0$ (si nuestras rectas son verticales), $\rho(c)$ es constante 0 correspondiendo al hecho de que $\{\mathbf{u}_c^\perp\}$ son las ortogonales a $\mathbf{e}_2^\perp$. Supongamos que $b > 0$ (el caso análogo $b < 0$ se deja como ejercicio). Entonces $\rho(c)$ tiene un mínimo, b, en $c = 0$, es simétrica ($\rho(c) = \rho(-c)$) y cumple que $\rho(c) \leq 1$ para $-1 \leq c \leq 1$, alcanzándose la igualdad en los extremos. Esto último corresponde a que $\mathbf{u}_c^\perp$ es concurrente a $\mathbf{e}_2^\perp$ cuando $|c| < 1$, y cuando $c = \pm 1$ son paralelas (con $(\pm 1, 0)$ como punto común al infinito); para $|c| > 1$, $\rho(c) > 1$ y las rectas $\mathbf{u}_c^\perp$ y $\mathbf{e}_2^\perp$ son ultraparalelas. Entonces $\rho(c)$ cae en el dominio de la función arcocoseno sólo en los dos primeros casos. En el primero, se aplica la fórmula (8.17), pero ésta **se puede extender continuamente** de pares de rectas concurrentes **para incluir pares de rectas paralelas**, lo que nos da que el ángulo de estas últimas es 0 (o π para el caso $b < 0$; ver Ejercicio 8.53).

Que $\rho(c)$ sea simétrica corresponde a que la rotación de π en el orígen lleva a $\mathbf{e}_2^\perp$ en sí misma y a $\mathbf{u}_c^\perp$ en $\mathbf{u}_{-c}^\perp$ invirtiendo orientación en ellas y consecuentemente tienen el mismo ángulo. Finalmente, que la función $\rho(c)$ baje de 1 a b mientras c va de -1 a 0 y suba simétricamente a 1 al seguir c de 0 a 1, implica que el ángulo $\measuredangle(\mathbf{u}_c^\perp, \mathbf{e}_2^\perp)$ empieza en 0 para $c = -1$, sube monótonamente hasta alcanzar su máximo, $\arccos(b) = \measuredangle(\mathbf{u}_0^\perp, \mathbf{e}_2^\perp)$, en $c = 0$ (que coincide con el ángulo euclidiano) y vuelve a bajar monotona y continuamente a 0 cuando $c = 1$. Como corolario de este hecho tenemos la siguiente proposición.

Proposicion 8.9 *Existe un polígono hiperbólico regular de* n *lados y ángulo interno* α *para cualquier* $n \geq 3$ *y* $0 < \alpha < \pi(n-2)/n$.

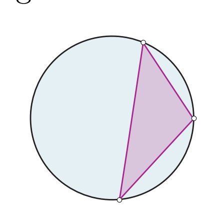
Demostración. Consideremos $(a, b) = (\cos(\pi/n), \operatorname{sen}(\pi/n))$ en la discusión anterior. Se tiene entonces que

$$\measuredangle(\mathbf{u}_0^\perp, \mathbf{e}_2^\perp) = \pi\frac{(n-2)}{2n}.$$



De nuestra última observación se sigue que existe una única c con $0 < c < 1$, para la cual $\measuredangle(\mathbf{u}_c^\perp, \mathbf{e}_2^\perp) = \alpha/2$. Sea P_c el polígono regular euclidiano de n lados con centro en el origen y vértice en $(c, 0)$; nótese que está contenido en el plano hiperbólico. Su grupo de simetrías es el grupo diédrico de orden n, $\mathbf{D}_n$, generado por las reflexiones en $\mathbf{e}_2^\perp$ y la recta por el origen con dirección (a, b). Como $\mathbf{D}_n \subset O(2) \subset \mathcal{H}(2)$, este polígono, considerado en el plano hiperbólico (puntos y segmentos de rectas ordenados cíclicamente), tiene todas las simetrías hiperbólicas posibles, es decir, cualquier par incidente "vértice-lado" se puede mandar en cualquier otro, y eso es lo que llamamos *regular*. Entonces todos sus ángulos internos son iguales y, por construcción, son iguales a α. □

Un caso importante es el límite cuando el ángulo α de los triángulos regulares tiende a 0; se les llama *triángulos hiperbólicos ideales.* Están formados por tres líneas hiperbólicas 2 a 2 paralelas, pierden sus vértices pues quedan en el infinito, y por el Teorema de 3 en 3 no importa qué tres puntos se tomen en $\mathbb{S}^1$; son equivalentes. Resulta que su área es π (con la noción natural de distancia que ahí viene) y es la máxima área que puede tener un triángulo hiperbólico.



EJERCICIO 8.50 Encuentra el vértice $(c, 0)$ de un cuadrado hiperbólico regular centrado en $(0, 0)$ con ángulo interior $\pi/3$, y el de un hexágono con ángulo $\pi/2$.

EJERCICIO 8.51 Define un triángulo isósceles como un triángulo con dos ángulos iguales. Demuestra que la suma de ángulos de un triángulo hiperbólico isósceles es menor que π.

EJERCICIO 8.52 Considera el ánalisis del ángulo entre $\mathbf{e}_2^\perp$ y $\mathbf{u}_c^\perp$ que dejamos pendiente, es decir, cuando $b < 0$. Demuestra que $\measuredangle(\mathbf{u}_c^\perp, \mathbf{e}_2^\perp)$ baja de π a $\measuredangle(\mathbf{u}_0^\perp, \mathbf{e}_2^\perp) = \arccos(b)$ conforme c va de -1 a 0 y luego vuelve a subir simétricamente.

EJERCICIO 8.53 Sean $\mathbf{u}$ y $\mathbf{v}$ dos vectores espaciales unitarios tales que $[\mathbf{u}] \neq [\mathbf{v}]$. Ya vimos que si sus líneas polares son concurrentes entonces $|\mathbf{u} \bullet_L \mathbf{v}|$ (en valor absoluto) es menor que 1. Demuestra que sus líneas polares son paralelas o ultraparalelas correspondiendo respectivamente a que $|\mathbf{u} \bullet_L \mathbf{v}|$ sea igual a 1 o mayor que 1.

8.4.2 Distancias

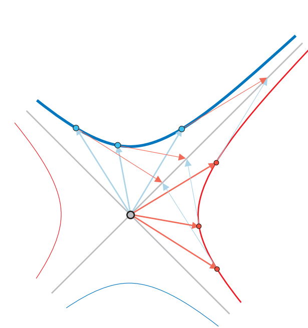
Por fin vamos a deducir la noción de distancia. Ya sabemos trasladar en rectas pero aún no sabemos medir distancias en ellas. Claramente estas dos nociones deben corresponderse; dada una traslación, cada punto en la recta y su imagen deben estar a la misma distancia. Veremos que esto obliga (salvo una constante) a la noción de distancia. Como todas las rectas hiperbólicas son iguales pues ya tenemos el grupo, bastará concentrarnos en una que pasa por el "origen" $\mathbf{e}_3$ para luego propagar la noción. Las rectas por $\mathbf{e}_3$ son aún "más iguales", se ven (pensaremos, ahora sí, en la cazuela hiperbólica) como una rama de una hipérbola canónica en $\mathbb{R}^2$, así que simplifiquemos las cosa y estudiemos esta última.

Lorentz en dimensión 2

Hasta nuevo aviso, denotemos con L a la análoga de la matriz de Lorentz en dimensión 2, que podríamos llamar "la Lorentzita"

$$L = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Usemos $\mathbf{u}\bullet_L \mathbf{v} = \mathbf{u}^\top L\mathbf{v}$ y $|\mathbf{u}|_L^2 = \mathbf{u}\bullet_L \mathbf{u}$ para denotar, respectivamente, a la forma bilineal y a la forma cuadrática asociada $(x^2 - y^2)$; donde ahora $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$. Igualmente, usaremos toda la terminología que desarrolamos para el caso de dimensión 3. Una gran diferencia es que ahora hay una correspondencia entre los vectores *unitarios temporales* y *espaciales*; ambos son hiperbólas cuadradas; la única diferencia es que una se abre hacia el eje y (la temporal) y la otra es "horizontal" (la espacial). Tenemos entonces una clara noción de *compadre* L-*ortogonal* (análoga a la euclidiana) pues si declaramos, cambiando la notación para que no se confunda con el ortogonal euclidiano,



$$(x, y)^L = (y, x)\,,$$

claramente se tiene que $\mathbf{u} \bullet_L \mathbf{u}^L = 0$ y $|\mathbf{u}|_L^2 = -\left|\mathbf{u}^L\right|_L^2$. Intercambia temporales por espaciales dejándoles la misma magnitud en valor absoluto. Los vectores luz forman dos rectas, una (la $x = y$) donde el compadre L-ortogonal es el mismo vector, y la otra (la $x = -y$) donde es su inverso. El compadre L-ortogonal tiene un elegante significado geométrico (análogo también al euclidiano respecto a los círculos).

Lema 8.10 *Dado* $\mathbf{u} \in \mathbb{R}^2, \mathbf{u} \neq 0$, *entonces* $\mathbf{u}^L$ *es tangente en* $\mathbf{u}$ *a la curva de nivel de la función* $|\mathbf{x}|_L^2$ *que pasa por* $\mathbf{u}$, *es decir, a la curva* $C_\mathbf{u} := \left\{\mathbf{x} \in \mathbb{R}^2 \mid |\mathbf{x}|_L^2 = |\mathbf{u}|_L^2\right\}$.

...Dibujo

Demostración. Obsérvese primero que las curvas de nivel $C_\mathbf{u}$ están definidas por una ecuación $x^2 - y^2 = cte$, y que por tanto son cónicas de las que sabemos hablar de sus tangentes. En el caso en que $\mathbf{u}$ es un vector luz (la cónica degenera en dos rectas) el lema es claro por la descripción de $\mathbf{u}^L$.

Supongamos entonces que $|\mathbf{u}|_L^2 \neq 0$. Para ver cómo intersecta la recta $\left[\mathbf{u}^L\right] + \mathbf{u}$ a la curva de nivel, tomemos un vector $t\,\mathbf{u}^L$ trasladado a $\mathbf{u}$ y evaluemos la forma cuadrática:

$$\begin{aligned}
\left|t\,\mathbf{u}^L + \mathbf{u}\right|_L^2 &= \left(t\,\mathbf{u}^L + \mathbf{u}\right) \bullet_L \left(t\,\mathbf{u}^L + \mathbf{u}\right) \\
&= \left(t\,\mathbf{u}^L\right) \bullet_L \left(t\,\mathbf{u}^L\right) + \left(t\,\mathbf{u}^L\right) \bullet_L \mathbf{u} + \mathbf{u} \bullet_L t\,\mathbf{u}^L + \mathbf{u} \bullet_L \mathbf{u} \\
&= t^2 \left(\mathbf{u}^L \bullet_L \mathbf{u}^L\right) + 2t\left(\mathbf{u}^\perp \bullet_L \mathbf{u}\right) + |\mathbf{u}|_L^2 \\
&= t^2 \left|\mathbf{u}^L\right|_L^2 + |\mathbf{u}|_L^2\,.
\end{aligned}$$

Puesto que $|\mathbf{u}|_L^2 \neq 0$ entonces $\left|\mathbf{u}^L\right|_L^2 \neq 0$, y de la ecuación anterior se deduce que $\left|t\,\mathbf{u}^L + \mathbf{u}\right|_L^2 = |\mathbf{u}|_L^2$, si y sólo si $t = 0$. Por tanto, la recta $\left[\mathbf{u}^L\right] + \mathbf{u}$ intersecta a $C_\mathbf{u}$ justo en $\mathbf{u}$ y es entonces su tangente ahí. □

Estudiemos ahora la rama positiva de temporales unitarios, que podemos llamar **la** "recta hiperbólica"

$$\mathbb{H}^1 := \left\{\mathbf{x} \in \mathbb{R}^2 \mid |\mathbf{x}|_L^2 = -1\,,\ \mathbf{x} \cdot \mathbf{e}_2 > 0\right\}.$$

Como en el caso de dimensión 3, tenemos el grupo $O(1,1)$ de matrices que dejan invariante la forma bilineal de Lorentz (o bien, tales que $B^\top LB = L$ o, equivalentemente, que sus columnas forman una base L-ortonormal). Dentro de éste, tenemos el subgrupo que manda a $\mathbb{H}^1$ en sí mismo, que llamamos $\mathcal{H}(1)$, y dentro de éste el de aquellas que preservan la orientación, $\mathcal{H}^+(1)$. Siguiendo los pasos del caso de dimensión 3, no es dificil ver que este último grupo es

$$\mathcal{H}^+(1) = \left\{ \begin{pmatrix} b & a \\ a & b \end{pmatrix} \mid a^2 - b^2 = -1 \text{ , } b > 0 \right\},$$

(es fácil comprobar que estas matrices cumplen que $B^\top LB = L$). Para $\mathbf{u} = (a, b) \in \mathbb{H}^1$, denotemos

$$T_\mathbf{u} = T_{(a,b)} := \begin{pmatrix} b & a \\ a & b \end{pmatrix}, \tag{8.18}$$

que es la *traslación* que manda al *origen* de la recta hiperbólica, $\mathbf{e}_2$, en $\mathbf{u}$, y desplaza a todo $\mathbb{H}^1$ dentro de sí mismo. Esta transformación líneal en $\mathbb{R}^2$ "aplasta" un cuadrante (en el que está $\mathbf{u}$) hacia su recta luz, o, mejor dicho, lo "alarga" pues preserva área, y compensa con un acortamiento de la otra recta luz; sus vectores propios son los luz con valores propios $a+b$ y $b-a$.

Obsérvese que (8.18) establece una biyección $\mathbb{H}^1 \to \mathcal{H}^+(1)$ que a cada punto le asocia la traslación que manda al origen en él. Nuestro problema actual es encontrar una distancia en $\mathbb{H}^1$ que logre que estas traslaciones sean rígidas.

Puesto que $\mathbb{H}^1$ es topológicamente una recta, encontrar una distancia en $\mathbb{H}^1$ compatible con estas traslaciones equivale a parametrizar adecuadamente $\mathbb{H}^1$. Es decir, a encontrar una función

$$\mathbf{h} : \mathbb{R} \to \mathbb{H}^1$$

que convierta a la suma (las traslaciones en $\mathbb{R}$, a partir de las cuales ahí se define distancia) en nuestras traslaciones hiperbólicas. Es decir, tal que al componerla con la biyección $\mathbb{H}^1 \to \mathcal{H}^+(1)$ mande a la suma en composición, que se escribe

$$T_{\mathbf{h}(t+r)} = T_{\mathbf{h}(t)} T_{\mathbf{h}(r)}. \tag{8.19}$$

Evaluando en el origen, la parametrización $\mathbf{h}$ que andamos buscando debe cumplir entonces que

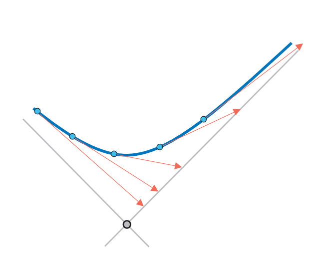
$$\mathbf{h}(t+r) = T_{\mathbf{h}(t)}(\mathbf{h}(r)), \tag{8.20}$$

pero además, como las traslaciones de $\mathcal{H}^+(1)$ quedan determinadas por su valor en el origen, esta condición es equivalente a la de *homomorfismo* (8.19). En particular, como la traslación por 0 es la identidad en $\mathbb{R}$, y $T_{\mathbf{e}_2}$ es la identidad en $\mathbb{H}^1$, $\mathbf{h}$ también debe cumplir la *condición inicial*:

$$\mathbf{h}(0) = \mathbf{e}_2. \tag{8.21}$$

Antes de deducir formalmente cuál es la parametrización $\mathbf{h}$ que cumple estas condiciones, vale la pena preguntarse ¿qué es lo más bonito que puede pasar? Pues con inusitada frecuencia en matemáticas la respuesta a esta última pregunta es la correcta, o al menos es premonitoria. Manténgase también en mente la analogía de que medir ángulos equivale a parametrizar adecuadamente a $\mathbb{S}^1$.

Hemos visto (Lema 8.10) que para cada punto $\mathbf{u}$ de $\mathbb{H}^1$, su compadre L-ortogonal $\mathbf{u}^L$ es un vector tangente a $\mathbb{H}^1$. Entonces $\mathbb{H}^1$ viene dado junto con un *campo de vectores* natural, es decir, en cada punto $\mathbf{u} \in \mathbb{H}^1$ hay un vector tangente distinguido, $\mathbf{u}^L$. Un punto, o partícula, viaja en $\mathbb{H}^1$ acorde con este campo vectorial si en cada momento su velocidad coincide con el vector tangente distinguido. Como estos vectores varían continuamente, es intuitivamente claro que se puede viajar de esta manera; el campo vectorial de compadritos define cómo deben moverse internamente los puntos. Si llamamos $\mathbf{h}$ a la función de posición de una partícula que viaja acorde con el campo, y la ponemos en el origen en el tiempo 0, *i.e.*, $\mathbf{h}(0) = \mathbf{e}_2$, obtenemos al candidato ideal para la parametrización. Pues obsérvese que cada punto sabe dejarse llevar por el *flujo,* así que si los ponemos a todos juntos y "agarrados" de la mano, se mueven coherentemente dando lugar a traslaciones. Entonces la condición de homomorfismo (8.19) equivale a dejar correr a $\mathbb{H}^1$ con el flujo por un cierto tiempo r, ahí tomar una foto instantánea ($T_{\mathbf{h}(r)}$: dónde anda cada uno), y luego dejarlo correr otro rato t; al final, simplemente se le dejó correr un tiempo $t + r$. Pero ¿por qué coinciden estos movimientos de acuerdo con el flujo con nuestras traslaciones hiperbólicas? Pues, como veremos, porque estas ultimas preservan el campo vectorial ya que preservan al L-compadrazgo.



Más formalmente, lo que le estamos pidiendo a la parametrización $\mathbf{h} : \mathbb{R} \to \mathbb{H}^1$ dada por el flujo es que su derivada sea el compadre L-ortogonal. Si expresamos $\mathbf{h}$ en coordenadas, $\mathbf{h} = (h_1, h_2)$, donde h_1 y h_2 son funciones de $\mathbb{R}$ en $\mathbb{R}$, entonces tenemos una condición inicial y una diferencial

$$\begin{aligned} \mathbf{h}(0) = \mathbf{e}_2 \quad &\Leftrightarrow \quad (h_1(0), h_2(0)) = (0, 1); \\ \mathbf{h}' = \mathbf{h}^L \quad &\Leftrightarrow \quad (h_1', h_2') = (h_2, h_1). \end{aligned} \tag{8.22}$$

Aunque la ecuación diferencial intercambia índices, de ella se sigue que $h_i'' = h_i$; es decir, ambas son funciones cuya aceleración es su valor: cuanto más lejos (positivo) del 0 estén, más aceleran y agarran una velocidad "endemoniada". Por ejemplo, de h_1 sabemos que $h_1(0) = 0$, que $h_1'(0) = 1$ y que $h_1'' = h_1$. No es este el libro para hacerlo en detalle, pero heurísticamente si determinamos la posición y la velocidad inicial y damos una regla para la aceleración de acuerdo con la posición, la función debe estar determinada de manera única. Si suponemos esto (que se sigue del Teorema de unicidad y existencia de soluciones de ecuaciones diferenciales —y que al manejar un coche con el simple acelerador se corrobora todos los días—), ya tenemos candidato para la parametrización $\mathbf{h}$; sus coordenadas tienen nombres clásicos: h_1 es el *seno hiperbólico* y h_2 el *coseno hiperbólico,* que se denotan senh y cosh respectivamente.

Aún suponiendo que ya tenemos $\mathbf{h} = (\text{senh}, \cosh)$ que cumple (8.22), nos falta demostrar más formalmente que cumple (8.20); si se lo cree ya el lector, sáltese la siguiente subsección que tiene que ver más con cálculo.

*Los cálculos

Queremos probar que la condición de homomorfismo (8.20) es equivalente a la condicion del "flujo de compadrazgo" (8.22), bueno, casi. Supongamos entonces que $\mathbf{h} : \mathbb{R} \to \mathbb{H}^1$ cumple que $\mathbf{h}(0) = \boldsymbol{e}_2$ que es la parte común. Veamos primero que (8.20) implica algo un poco más general que (8.22).

En coordenadas (8.20) se expresa, usando (8.18), como

$$\begin{aligned}\begin{pmatrix} h_1(t+r) \\ h_2(t+r) \end{pmatrix} &= \begin{pmatrix} h_2(t) & h_1(t) \\ h_1(t) & h_2(t) \end{pmatrix}\begin{pmatrix} h_1(r) \\ h_2(r) \end{pmatrix} \\ &= \begin{pmatrix} h_2(t)\,h_1(r) + h_1(t)\,h_2(r) \\ h_1(t)\,h_1(r) + h_2(t)\,h_2(r) \end{pmatrix}.\end{aligned}$$

Que son las identidades aditivas para el seno y el coseno hiperbólicos.

De estas dos ecuaciones podemos sacar condiciones diferenciales que deben cumplir las funciones. Consideremos la segunda, y con base en su identidad para la suma expresemos su derivada en la forma clásica de límites:

$$\begin{aligned} h_2'(t) &= \lim_{r\to 0}\left(\frac{h_2(t+r) - h_2(t)}{r}\right) \\ &= \lim_{r\to 0}\left(\frac{h_1(t)\,h_1(r) + h_2(t)\,h_2(r) - h_2(t)}{r}\right) \\ &= \lim_{r\to 0}\left(\frac{h_1(t)\,h_1(r) + h_2(t)\,(h_2(r) - 1)}{r}\right) \\ &= \lim_{r\to 0}\left(\frac{h_1(t)\,h_1(r)}{r}\right) + \lim_{r\to 0}\left(\frac{h_2(t)\,(h_2(r) - 1)}{r}\right) \\ &= h_1(t)\lim_{r\to 0}\left(\frac{h_1(r)}{r}\right) + h_2(t)\lim_{r\to 0}\left(\frac{h_2(r) - 1}{r}\right) \\ &= h_1(t)\lim_{r\to 0}\left(\frac{h_1(0+r) - h_1(0)}{r}\right) + h_2(t)\lim_{r\to 0}\left(\frac{h_2(0+r) - h_2(0)}{r}\right) \\ &= h_1(t)\; h_1'(0) + h_2(t)\; h_2'(0)\,. \end{aligned}$$

Pero además $\mathbf{h}$ va por la hipérbola $\mathbb{H}^1$ y pasa por su mínimo en el 0, lo cual implica que $h_2'(0) = 0$. Así que se cumple

$$h_2'(t) = h_1'(0)\; h_1(t)\,.$$

De un cálculo análogo (que dejamos como ejercicio) se obtiene que

$$h_1'(t) = h_1'(0)\; h_2(t)\,.$$

Éstas son, salvo la constante $h_1'(0)$ (el casi de la equivalencia), las ecuaciones (8.22). Esta constante correponde a viajar con velocidades distintas en el flujo, o bien, considerar múltiplos del campo vectorial. Sin embargo, si tomamos esta constante igual a 1 tenemos que la condición de homomorfismo (8.20) implica la del flujo (8.22).

Supongamos ahora que $\mathbf{h}' = \mathbf{h}^L$ (8.22), y vamos a demostrar que $\mathbf{h}(t+r) = T_{\mathbf{h}(t)}(\mathbf{h}(r))$ (8.20). La idea básica es que podemos reparametrizar y luego usar (es inevitable) la unicidad de soluciones.

Pensemos a t como una constante; es más, denotemos $\mathbf{u} = (a, b) = \mathbf{h}(t)$ y definamos la función $\mathbf{g} : \mathbb{R} \to \mathbb{H}^1$ como

$$\mathbf{g}(r) = T_{\mathbf{u}}^{-1}(\mathbf{h}(t+r)).$$

Claramente se cumple que $\mathbf{g}(0) = \mathbf{e}_2$. Además, derivando con respecto a r y como la matriz $T_{\mathbf{u}}^{-1}$ es constante, y multiplicar por ella es lineal, se cumple que

$$\mathbf{g}'(r) = T_{\mathbf{u}}^{-1}(\mathbf{h}'(t+r)) = T_{\mathbf{u}}^{-1}\left(\mathbf{h}(t+r)^L\right) = \left(T_{\mathbf{u}}^{-1}(\mathbf{h}(t+r))\right)^L = \mathbf{g}(r)^L,$$

pues $T_{\mathbf{u}}^{-1} \in \mathcal{H}^+(1)$ y es fácil ver que estas matrices preservan el compadre L-ortogonal. (Si alguien se siente incómodo con que hayamos derivado vectores, o que haya entrado la derivada tan fácil, se le invita a hacer las cuentas en coordenadas usando (8.18) y el Ejercicio 8.54). Entonces $\mathbf{g}$ cumple las mismas condiciones que $\mathbf{h}$ y por la unicidad de soluciones se tiene que son la misma parametrización de $\mathbb{H}^1$. Multiplicando ambos lados de la ecuación $\mathbf{h} = \mathbf{g}$ por $T_{\mathbf{u}}$, tenemos

$$T_{\mathbf{u}}(\mathbf{h}(r)) = T_{\mathbf{u}}(\mathbf{g}(r)) = T_{\mathbf{u}}\left(T_{\mathbf{u}}^{-1}(\mathbf{h}(t+r))\right) = \mathbf{h}(t+r),$$

que es lo que queríamos demostrar.

EJERCICIO 8.54 Demuestra que si $T_{\mathbf{u}}$ es como en (8.18) entonces

$$T_{\mathbf{u}}^{-1} = \begin{pmatrix} b & -a \\ -a & b \end{pmatrix}.$$

EJERCICIO 8.55 Demuestra que si $T \in \mathcal{H}^+(1)$ entonces $T\left(\mathbf{u}^L\right) = (T\mathbf{u})^L$.

EJERCICIO 8.56 Demuestra que si $\mathbf{h} : \mathbb{R} \to \mathbb{H}^1$ cumple la condición de homomorfismo (8.19), entonces $\mathbf{g}(t) = \mathbf{h}(c\,t)$ para cualquier constante $c \neq 0$, también la cumple.

Seno y coseno hiperbólicos (distancia en $\mathbb{H}^1$)

Resumiendo, tenemos que la función $\mathbf{h} = (\sinh, \cosh)$ es una biyección de $\mathbb{R}$ con $\mathbb{H}^1$ que hace corresponder a las traslaciones reales con las hiperbólicas, pues las funciones trigonométricas hiperbólicas, $\cosh t$ y $\sinh t$ (cuyas gráficas aparecen al margen) cumplen:

...Dibujo, graficas de cosh y senh

	$\cosh^2 t - \sinh^2 t = 1$	
$\cosh 0 = 1$		$\sinh 0 = 0$
$\cosh' t = \sinh t$		$\sinh' t = \cosh t$
	$\cosh(t+r) = \cosh t \cosh r + \sinh t \sinh r$	
	$\sinh(t+r) = \cosh t \sinh r + \sinh t \cosh r$	

La distancia en $\mathbb{H}^1$ debe ser

$$d(\mathbf{h}(t), \mathbf{h}(r)) = |t - r|.$$

Para expresarla sin referencia a $\mathbf{h}$, podemos usar la inversa del coseno hiperbólico, $\cosh^{-1} t$, que tiene sentido sólo para $t \geq 1$, que, por definición, siempre toma valores positivos, y que de la segunda coordenada de un punto en $\mathbb{H}^1$ nos dice, en valor absoluto, de qué t proviene. En efecto, para el origen $\mathbf{e}_2 = \mathbf{h}(0)$, se tiene que

$$d(\mathbf{e}_2, \mathbf{h}(t)) = |t| = \cosh^{-1}(\mathbf{e}_2 \cdot \mathbf{h}(t)) = \cosh^{-1}(-(\mathbf{e}_2 \bullet_L \mathbf{h}(t))).$$

Esta última expresión (como en los ángulos) es la base para generalizar, pues podemos definir entonces para cualquier $\mathbf{u}, \mathbf{v} \in \mathbb{H}^1$:

$$d(\mathbf{u}, \mathbf{v}) = \cosh^{-1}(-(\mathbf{u} \bullet_L \mathbf{v})).$$

Que es claramente invariante bajo $\mathcal{H}(1)$ y concuerda con la distancia al origen; hace que la parametrización $\mathbf{h}$ sea una isometría.

Distancia en $\mathbb{H}^2$

Regresando al plano hiperbólico $\mathbb{H}^2 \subset \mathbb{R}^3$, debe ser claro que la buena definición de *distancia* de $\mathbf{u}$ a $\mathbf{v}$ (puntos de $\mathbb{H}^2$) es

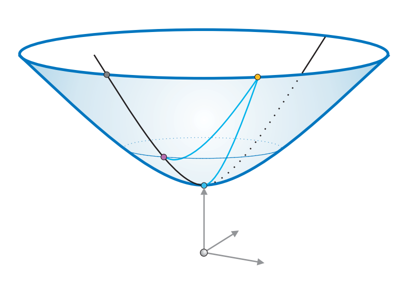
$$\boxed{d(\mathbf{u}, \mathbf{v}) = \cosh^{-1}(-(\mathbf{u} \bullet_L \mathbf{v})).} \tag{8.23}$$

Pues es claramente invariante bajo $\mathcal{H}(2)$ ya que estas transformaciones preservan la forma bilineal de Lorentz y entonces el resultado de la fórmula no cambia. Además, hace justo lo que queremos, pues dados cualesquiera $\mathbf{u}, \mathbf{v} \in \mathbb{H}^2$ podemos trasladar al origen con alguna $f \in \mathcal{H}(2)$ tal que $f(\mathbf{u}) = \mathbf{e}_3$, entonces

$$\begin{aligned} d(\mathbf{u}, \mathbf{v}) &= \cosh^{-1}(-(\mathbf{u} \bullet_L \mathbf{v})) \\ &= \cosh^{-1}(-(f(\mathbf{u}) \bullet_L f(\mathbf{v}))) \\ &= \cosh^{-1}(-(\mathbf{e}_3 \bullet_L f(\mathbf{v}))); \end{aligned}$$

y $-(\mathbf{e}_3 \bullet_L f(\mathbf{v}))$ es precisamente la tercera coordenada de $f(\mathbf{v})$, que, vista en el plano vertical correspondiente, igual a $\mathbb{R}^2$, es la segunda coordenada de un punto en $\mathbb{H}^1$.

Por está última expresión, como todos los puntos de $\mathbb{H}^2$ están arriba de $\mathbf{e}_3$, se tiene claramente que $d(\mathbf{u}, \mathbf{v}) = 0$ si y sólo si $\mathbf{u} = \mathbf{v}$; o bien, que $\mathbf{u} \bullet_L \mathbf{v} \leq -1$ y la igualdad se da si y sólo si $\mathbf{u} = \mathbf{v}$. Pero nos falta probar **la desigualdad del triángulo** para ver que hemos definido a una distancia con todas las de la ley.

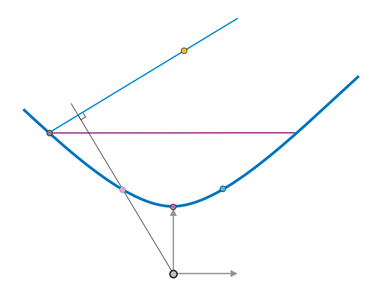


Sean $\mathbf{u}, \mathbf{v}, \mathbf{w}$ tres puntos en la cazuela hiperbólica $\mathbb{H}^2$; queremos demostrar que

$$d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w}) \geq d(\mathbf{u}, \mathbf{w}).$$

Podemos suponer que $\mathbf{u} = \mathbf{e}_3$ y además que $\mathbf{v}$ está en el plano yz, y entonces que se escribe $\mathbf{v} = (0, a, b)$. Sea c la *altura* de $\mathbf{w}$, es decir, $c = \mathbf{w} \cdot \mathbf{e}_3$ que es la que determina la distancia a $\mathbf{e}_3$. Si $\mathbf{v}$ estuviera más alto que $\mathbf{w}$ (o sea, si $b \geq c$) entonces $d(\mathbf{u}, \mathbf{v}) \geq d(\mathbf{u}, \mathbf{w})$ y ya no hay nada que demostrar; supongamos entonces que $\mathbf{v}$ está debajo de $\mathbf{w}$. Sea $\mathbf{v}'$ el punto en el rayo de $\mathbf{u}$ a $\mathbf{v}$ que está a la altura de $\mathbf{w}$; de tal manera que $d(\mathbf{u}, \mathbf{w}) = d(\mathbf{u}, \mathbf{v}')$, es decir, ambos están en un *círculo* hiperbólico centrado en $\mathbf{e}_3$, que es la intersección de $\mathbb{H}^2$ con el plano horizontal $z = c$. Nótese que los círculos horizontales $\Pi_r \cap \mathbb{H}^2$, donde $\Pi_r : z = r$, son los *círculos* en torno a $\mathbf{e}_3$, esto es, los conjuntos de puntos con una misma distancia a $\mathbf{e}_3$.

Sea $T \in \mathcal{H}^+(2)$ la traslación hiperbólica que lleva a $\mathbf{v}$ al origen hiperbólico $\mathbf{e}_3$. Es fácil ver que es la matriz



$$T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & b & -a \\ 0 & -a & b \end{pmatrix}$$

y que ésta lleva a los planos horizontales Π_r en la familia de planos $ay + bz = r$ normales al vector $(0, a, b)$. De aquí se sigue que la imagen del círculo $\Pi_c \cap \mathbb{H}^2$ es una elipse con un único punto de tercera coordenada mínima, a saber $T(\mathbf{v}')$. Por tanto, $T(\mathbf{w})$ está arriba de $T(\mathbf{v}')$ y se tiene

$$d(\mathbf{u}, \mathbf{w}) = d(\mathbf{u}, \mathbf{v}') = d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{v}') \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w}),$$

con la igualdad si y sólo si $\mathbf{u}, \mathbf{v}, \mathbf{w}$ son colineales (pues la igualdad se da si y sólo si $d(\mathbf{v}, \mathbf{v}') = d(\mathbf{v}, \mathbf{w})$, si y sólo si $T(\mathbf{w})$ coincide con el mínimo $T(\mathbf{v}')$).

Como corolario de la demostración, obsérvese que los círculos hiperbólicos alrededor de un punto cualquiera $\mathbf{u} \in \mathbb{H}^2$ son las intersecciones de la cazuela $\mathbb{H}^2$ con los planos paralelos a su plano tangente en $\mathbf{u}$; pues esto es cierto en $\mathbf{e}_3$.

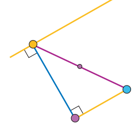
EJERCICIO 8.57 Demuestra que el conjunto de puntos que equidistan de dos puntos dados es un recta, llamada su *mediatriz.*

EJERCICIO 8.58 Recuerda que dos líneas ultraparalelas tienen una única perpendicular común, y observa además que si tienen una perpendicular común son ultraparalelas. Demuestra que los puntos de intersección de la perpendicular común con las rectas es donde se realiza la distancia mínima entre ellas (usa las reflexiones y la desigualdad del triángulo).

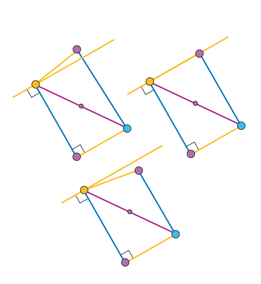
8.4.3 Triángulos

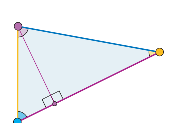
Queremos demostrar que **la suma de ángulos de un triángulo hiperbólico es menor que π, y** convencer que **su área es acotada**.

Un triángulo hiperbólico está determinado por tres puntos a, b, c (ya no será relevante si los pensamos en $\mathbb{P}^2$ o en $\mathbb{R}^3$; también los podemos pensar en abstracto), pues, como en el caso euclidiano, quedan determinados los segmentos que definen y sus ángulos internos, llamémosles α, β, γ respectivamente.

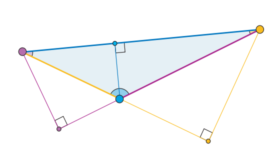


Consideremos primero el caso de un triángulo rectángulo: supongamos que $\gamma = \pi/2$. Sea ℓ la recta ortogonal a $\langle c, b\rangle$ en el punto b y observemos (Ejercicio 8.58) que el segmento bc realiza la distancia de la recta ℓ a la recta $\langle a, c\rangle$. Sea f la rotación de π en el punto medio del segmento ab. Se tiene que $f(a) = b$ y $f(b) = a$; queremos ver que $f(c)$ está en la franja entre ℓ y la recta $\langle a, c\rangle$. Si $f(c)$ estuviera del otro lado de ℓ que c entonces el segmento de $f(c)$ a $a = f(b)$ intersectaría a ℓ y tendríamos que la distancia de ℓ a la línea $\langle a, c\rangle$ es menor que $d(b, c)$ contradiciendo el Ejercicio 8.58. Y se contradice la unicidad de la perpendicular común si $f(c) \in \ell$; por tanto, $f(c)$ está del lado de a (y c) de la recta ℓ. Esto implica que $\alpha + \beta < \pi/2$ y por lo tanto que $\alpha + \beta + \gamma < \pi$.





Consideremos ahora cualquier triángulo. Tracemos la perpendicular a un lado desde el vértice opuesto, c digamos, que podemos llamar la *altura de* c. Hay dos casos por analizar; cuando la perpendicular corta en el segmento ab, y cuando corta a la línea $\langle a, b\rangle$ fuera de él. En el primer caso tenemos que el triángulo original se parte en dos triangulitos rectángulos. Como en cada uno de ellos la suma de sus dos ángulos no rectos es menor que $\pi/2$, la suma de los tres ángulos del grande da menos de π. El segundo caso se reduce al primero, pues al tomar la altura de los otros dos vértices, alguna de ellas parte el triángulo en dos. Esto completa la prueba de que $\alpha + \beta + \gamma < \pi$.

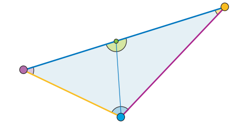


Al hacer con cuidado las cuentas de este último razonamiento aparece algo más profundo (o preciso) que conviene analizar. Dado un triángulo T con ángulos internos α, β, γ podemos definir su *déficit angular* como

$$\delta(T) = \pi - (\alpha + \beta + \gamma) .$$

Queríamos demostrar que el déficit angular de cualquier triángulo es positivo. Lo hicimos primero para triángulos rectángulos; y al extender este resultado descomponiendo un triángulo en dos triangulitos lo que se usa en el fondo es que el déficit angular es *aditivo*. Veámoslo en general.

Si tomamos cualquier punto de un lado de un triángulo T y trazamos el segmento al vértice opuesto, c digamos, se obtienen dos triangulitos T_1 y T_2 con ángulos complementarios nuevos θ_1 y θ_2, de tal manera que



$$\begin{aligned}\delta(T_1)+\delta(T_2) &= \pi-(\alpha+\theta_1+\gamma_1)+\pi-(\beta+\theta_2+\gamma_2)\\ &= 2\pi-(\alpha+\beta+(\gamma_1+\gamma_2)+(\theta_1+\theta_2))\\ &= 2\pi-(\alpha+\beta+\gamma+\pi)=\pi-(\alpha+\beta+\gamma)=\delta(T)\,.\end{aligned}$$

O bien, si tomamos un punto en el interior de T y *subdividimos* en tres triangulitos, se obtiene que

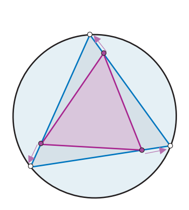
$$\delta(T_1)+\delta(T_2)+\delta(T_3)=\delta(T)\,,$$

pues los ángulos alrededor del nuevo vértice suman 2π, y el resto de los ángulos se asocia por parejas para dar los del original.

Así que si subdividimos un triángulo T en triangulitos que sólo se intersecten en sus fronteras (una iteración de las dos subdivisiones que acabamos de analizar) obtenemos que $\delta(T)$ es la suma de los déficit angulares de sus triangulitos:

¡La función $\delta(T)$ se comporta como una función de área!

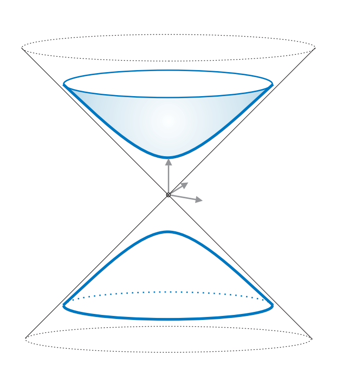
Pues medir el área de algo equivale a partir en pedazos chicos que no se traslapen y sumar sus áreas. Es concebible además, definir áreas no sólo de triángulos sino de cualquier figura hiperbólica con un proceso de triángulación, medir en triangulitos, y sumar; y en todo caso, un proceso de límites para figuras cuya frontera no esté segmentada. Ya no tenemos tiempo, espacio o aliento para llevar a cabo este programa, pero debe ser cierto. Y además que cualquier noción coherente de área debe ser un múltiplo de ésta que se nos atravesó.



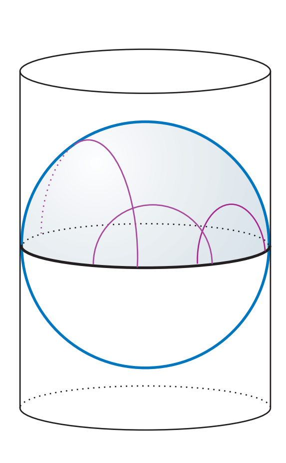
Se tiene que entre más chico sea un triángulo, más se parece a uno euclidiano; sus ángulos casi suman π y su déficit angular o área es casi cero. En el otro extremo, el área de cualquier triángulo está acotada por el área de los triángulos ideales (con sus tres vértices al infinito) que, como tienen ángulos de 0 grados, tienen deficit angular π. Estos se pueden ver como el límite de triángulos cuya área $\delta(T)$ es creciente pero tiende a π.

8.5 Modelos de Poincaré y el hemiplano superior

Además de los dos modelos del plano hiperbólico con que hemos trabajado hay otros dos importantes que describiremos en esta sección.

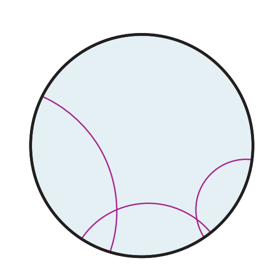


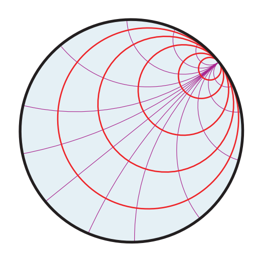
Consideremos al plano hiperbólico $\mathbb{H}^2$ como la cazuela contenida en $\mathbb{R}^3$ pero completando a éste último a $\mathbb{P}^3$, es decir, añadiendo un plano proyectivo al infinito. Formalmente, tomamos una nueva coordenada homogénea, w digamos, e identificamos (x, y, z) con $[x : y : z : 1]$. Nos interesa ver a $\mathbb{H}^2$ alrededor del infinito. Para esto podemos traer al plano al infinito al plano horizontal xy intercambiando las dos últimas coordenadas, $[x : y : z : w] \mapsto [x : y : w : z]$ ($\mapsto (x/z, y/z, w/z)$ en $\mathbb{R}^3$) que intercambia el plano $z = 0$ con el plano al infinito $w = 0$ (o equivalentemente, ver las cosas en la carta coordenada $z \neq 0$). El efecto en la cazuela hiperbólica de esta transformación (vista en $\mathbb{R}^3$) es mandarla en el hemisferio superior de la esfera unitaria $\mathbb{S}^2$; los unitarios temporales de abajo (con z negativa) van en el hemisferio inferior, y el ecuador es la imagen del cono de luz intersectado con el plano al infinito, donde se pegan ambas hojas. Pues los unitarios temporales están dados por la ecuación $x^2 + y^2 - z^2 = -1$, cuyo polinomio homogéneo asociado es $x^2 + y^2 + w^2 - z^2$, que al tomar $z = 1$ (y mandar el plano $z = 0$ al infinito) nos da la ecuación $x^2 + y^2 + w^2 = 1$ en la carta coordenada afín $z \neq 0$ parametrizada por xyw. Más aún. El cono de luz (con polinomio homogéneo $x^2+y^2-z^2$) se convierte en el cilindro circular $x^2+y^2 = 1$; efectivamente, todas sus rectas pasan ahora por el infinito vertical $[0 : 0 : 1 : 0]$; y los planos que pasaban por el origen son ahora planos verticales.



Tenemos entonces el **modelo esférico de** $\mathbb{H}^2$ que consiste en el hemisferio superior de $\mathbb{S}^2$ (todos sus puntos de tercera coordenada estrictamente positiva) con el círculo al infinito en el ecuador, y las líneas hiperbólicas son intersecciones con planos verticales, es decir semicírculos verticales, o bien, que intersectan ortogonalmente al ecuador. Notemos que el plano $z = 1$ (o, proyectivamente $z = w$) se ha quedado en su lugar. Así que la identificación del modelo proyectivo con el modelo esférico es simplemente proyectar al interior del círculo en el plano $z = 1$ ortogonalmente sobre el hemisferio superior de la esfera $\mathbb{S}^2$.

El **modelo de Poincaré** se obtiene del modelo esférico proyectando estereográficamente desde el polo sur $(0, 0, -1)$ sobre el plano xy del ecuador. Se tiene entonces que como conjunto de puntos es de nuevo el interior del círculo unitario, pero ahora las rectas hiperbólicas son círculos ortogonales a la frontera. Los puntos se mueven hacia el centro de tal manera que lo que antes eran segmentos de recta se vuelven ahora segmentos de círculos que llegan perpendicularmente a la frontera. Lo importante de este modelo es que es *conforme*, es decir, los ángulos hiperbólicos coinciden con los ángulos euclidianos en que se encuentran las líneas (que ahora son círculos, insistimos). Este modelo es el que usa Escher en sus grabados famosos.





Otra cosa que hay que notar del modelo de Poincaré es que en él los horociclos (recuérdese, las curvas que son ortogonales a un haz paralelo) son de nuevo círculos. Precisamente los circulitos tangentes a $\mathbb{S}^1$ en el punto al infinito correspondiente al haz.

El **modelo del hemiplano superior** se obtiene también del modelo esférico por una proyección estereográfica, pero ahora desde un punto en el ecuador (en el infinito) sobre una pantalla vertical que se identifica con el plano de los números complejos, $\mathbb{C}$. Como el círculo al infinito va a dar a una recta, a ésta se le identifica con la recta real en $\mathbb{C}$ de tal manera que la imagen de $\mathbb{H}^2$ resulta ser los números complejos cuya parte imaginaria es positiva; y de ahí su nombre. De nuevo, este modelo es conforme. Las rectas hiperbólicas aquí son semicírculos ortogonales a la recta real y líneas verticales. Los horocíclos son círculos tangentes a la recta real y líneas horizontales, que corresponden al único punto al infinito que no vemos (el infinito real); en estos horocíclos, la métrica inducida es euclidiana, así que la distancia se vuelve manejable, es inversamente proporcional a la altura. Además, lo bonito de este modelo es que la identificación del grupo hiperbólico $\mathcal{H}^+(2)$ con las transformaciones de Moebius se hace nítida. A cada transformación de Moebius se le puede asociar de manera única una matriz real



$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

con determinante 1. Y la fórmula

$$\mu(x) = \frac{ax + b}{cx + d}$$

se extiende tal cual a una acción sobre el hemiplano superior que es justo la correspondiente transformación hiperbólica. Queda entonces el plano hiperbólico dentro de la esfera de Riemann (ver página 252).

8.6 *Subgrupos discretos

La riqueza exhuberante de las simetrías en el plano hiperbólico sigue explorándose hoy día, así que no podemos ser exhaustivos; pero sí podemos señalar dos hechos importantes que indican lo profusos que son los subgrupos discretos de $\mathcal{H}(2)$.

Notemos primero que la Proposición 8.9 nos dice con cuáles polígonos regulares como losetas se puede armar un mosaico que cubra el plano hiperbólico. Para que m polígonos regulares de n lados se ajusten alrededor de un vértice se necesita que su ángulo interno sea $2\pi/m$, y por la Proposición 8.9 tal polígono existe si y sólo si

$$\begin{aligned} \frac{2\pi}{m} &< \frac{\pi(n-2)}{n} \\ \frac{2}{m} &< 1-\frac{2}{n} \\ \frac{1}{n}+\frac{1}{m} &< \frac{1}{2}. \end{aligned}$$

Así que para cada pareja n, m que cumpla la desigualdad anterior, se pueden pegar y pegar n-ágonos regulares con ángulo interno $2\pi/m$ por sus aristas de tal manera que ajusten m alrededor de cada vértice. Al mosaico resultante, que cubre $\mathbb{H}^2$, lo podemos llamar $\{\!\{n, m\}\!\}$ usando la notación de Schlafli que introdujimos al final del capítulo **??**. Nótese que los ocho valores de n y m que no cumplen la desigualdad anterior (y con $n, m \geq 3$ para que tenga sentido) dan lugar a los mosaicos euclidianos (el $\{\!\{3,6\}\!\}$, el $\{\!\{4,4\}\!\}$ y el $\{\!\{6,3\}\!\}$) y a los esféricos o platónicos (el $\{\!\{3,3\}\!\}$, el $\{\!\{3,4\}\!\}$, el $\{\!\{4,3\}\!\}$, el $\{\!\{3,5\}\!\}$ y el $\{\!\{5,3\}\!\}$). Dibujo...

El grupo de simetrías del mosaico hiperbólico $\{\!\{n, m\}\!\}$, esto es, las isometrías hiperbólicas que preservan la colección de polígonos $\{\!\{n, m\}\!\}$, está generado por tres reflexiones. Al escoger uno de los n-ágonos tenemos las dos reflexiones ρ_0, ρ_1 que generan el diédrico $\mathbf{D}_n$ de sus simetrías, y además está la reflexión ρ_2 en su arista que lo manda a su polígono adyacente (ρ_1 y ρ_2 generan entonces al diédrico $\mathbf{D}_m$ de simetrías que fijan el vértice base). Se tiene entonces que ρ_0, ρ_1, ρ_2 son las reflexiones en los lados de un triángulo con ángulos internos π/n, π/m, $\pi/2$. Llamemos $\Gamma_{n,m,2}$ a este grupo.

Claramente se puede generalizar a los llamados *grupos triangulares* $\Gamma_{n,m,k}$ (o caleidoscópicos) generados por las reflexiones ρ_0, ρ_1, ρ_2 en los lados de un triángulo con ángulos internos π/n, π/m, π/k. Entonces los generadores cumplen las relaciones

$$\begin{gathered} \rho_0^2 = \rho_1^2 = \rho_2^2 = \mathbf{1} \\ (\rho_0\rho_1)^n = (\rho_1\rho_2)^m = (\rho_2\rho_0)^k = \mathbf{1}, \end{gathered}$$

donde $\mathbf{1}$ es la identidad, y estas relaciones definen en abstracto al grupo. Cuando

$$\frac{1}{n}+\frac{1}{m}+\frac{1}{k} < 1$$

este grupo se representa geométricamente en el plano hiperbólico, es decir $\Gamma_{n,m,k} \subset \mathcal{H}(2)$; cuando la misma suma da 1 es de isometrías euclidianas y lo habíamos visto en la Sección 3.9.2, y cuando la suma excede a 1 es esférico (ver Sección 5.4.3).

Nótese, además, que los grupos $\Gamma_{n,m,k}$ hiperbólicos cumplen las dos propiedades de los que habíamos llamado grupos cristalográficos en 3.9.2 al cambiar $\mathbb{R}^2$ por $\mathbb{H}^2$; es decir, son discretos y "*cocompactos*".

Capítulo 9

Cónicas IV (tangentes y polaridad)

En este último capítulo recopilamos brevemente las implicaciones que nuestro desarrollo de la geometría hiperbólica tiene para las cónicas en general. Para construir la geometría hiperbólica usamos la forma bilineal de Lorentz asociada al círculo unitario, pero se puede hacer lo análogo para cualquier cónica.

9.1 Forma bilineal de una cónica

Consideremos una cónica $\mathcal{C}$ en $\mathbb{R}^2$ definida por un polinomio cuadrático $P(x, y)$. Hemos visto (capítulo 7) que $\mathcal{C}$ tiene una completación proyectiva (obtenida al homogeneizar el polinomio) que seguiremos llamando $\mathcal{C} \subset \mathbb{P}^2$ (al fin que se le añaden a lo más dos puntos). Y que una cónica proyectiva tiene una matriz simétrica A (de 3×3) asociada, de tal manera que su polinomio homogéneo, que también hemos llamado su *forma cuadrática* asociada, se escribe

$$\tilde{P}(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} = \mathbf{x} \cdot (A\mathbf{x}) .$$

Tiene entonces una *forma bilineal simétrica* asociada, al diferenciar las dos apariciones de la variable $\mathbf{x}$:

$$\mathbf{x} \bullet_A \mathbf{y} := \mathbf{x}^\top A\mathbf{y} = \mathbf{x} \cdot (A\mathbf{y}) = (A\mathbf{x}) \cdot \mathbf{y}. \tag{9.1}$$

Ésta es una función de dos variables en $\mathbb{R}^3$ que toma valores reales. Y cumple claramente que

$$\begin{aligned} \mathbf{x} \bullet_A \mathbf{y} &= \mathbf{y} \bullet_A \mathbf{x} && \text{(es simétrica);} \\ (\alpha\mathbf{x} + \beta\,\mathbf{y}) \bullet_A \mathbf{z} &= \alpha\,(\mathbf{x} \bullet_A \mathbf{z}) + \beta\,(\mathbf{y} \bullet_A \mathbf{z})\,; && \\ \mathbf{x} \bullet_A (\alpha\,\mathbf{y} + \beta\,\mathbf{z}) &= \alpha\,(\mathbf{x} \bullet_A \mathbf{y}) + \beta\,(\mathbf{x} \bullet_A \mathbf{z}) && \text{(es bilineal).} \end{aligned}$$

Las dos últimas ecuaciones son redundantes; basta con una de ellas pues entonces de la simetría se obtiene la otra; las ponemos simplemente para enfatizar su carácter *bilineal* (lineal en cada variable).

9.2 Tangentes y polaridad

Supongamos que $\det A \neq 0$, es decir, que la cónica original $\mathcal{C}$ es no degenerada. Podemos definir entonces para cualquier $[\mathbf{u}] \in \mathbb{P}^2$, su recta A-*polar*, que denotamos $[\mathbf{u}]^A$ por la ecuación lineal homogénea

$$[\mathbf{u}]^A : \quad \mathbf{u} \bullet_A \mathbf{x} = (A\mathbf{u}) \cdot \mathbf{x} = 0.$$

Para demostrar que este polinomio lineal homogéneo no es cero y que $[\mathbf{u}]^A$ es efectivamente una línea, es necesario suponer que $\det A \neq 0$; pues entonces, como $\mathbf{u} \neq \mathbf{0}$ se tiene que $A\mathbf{u} \neq \mathbf{0}$.

Decimos que dos puntos $[\mathbf{u}], [\mathbf{v}] \in \mathbb{P}^2$ son A-*polares* o A-*ortogonales* si $\mathbf{u} \bullet_A \mathbf{v} = 0$. Es claro que es una relación simétrica. En particular

$$[\mathbf{u}] \in [\mathbf{v}]^A \qquad \Leftrightarrow \qquad [\mathbf{v}] \in [\mathbf{u}]^A . \tag{9.2}$$

Si todos nuestros datos estan en $\mathbb{R}^2$, es decir, si $[\mathbf{u}] = [a : b : 1]$, entonces la ecuación de $[\mathbf{u}]^A$ tiene la forma

$$(a, b, 1)\, A \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \left(A \begin{pmatrix} a \\ b \\ 1 \end{pmatrix} \right) \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0,$$

que es una línea en $\mathbb{R}^2$ siempre y cuando $[A\mathbf{u}] \neq [0 : 0 : 1]$. Si $\mathbf{u}$ es tal que $[A\mathbf{u}] = [0 : 0 : 1]$ entonces su línea A-polar es la línea al infinito y $[\mathbf{u}]$ es el *centro* de la cónica; de aquí que el centro, que habíamos detectado algebraicamente, es sólo un invariante afín, y que las parábolas "no tienen" centro: también está en el infinito.

Regresando a $\mathbb{P}^2$, para que no aparezcan casos especiales, tenemos que toda línea ℓ es la A-polar de un punto. Pues si ℓ está dada por la ecuación $\mathbf{v} \cdot \mathbf{x} = 0$, como A es invertible, tomamos $\mathbf{u} = A^{-1}\mathbf{v}$ y entonces ℓ es A-polar de $[\mathbf{u}]$. Así que también podemos hablar del *punto* A-*polar* de una línea y **la A-polaridad establece una biyección entre puntos y líneas de $\mathbb{P}^2$**.

Hemos estudiado dos casos. Primero, cuando A es la identidad: la I-polaridad corresponde a la ortogonalidad euclidiana o la *dualidad* en el proyectivo, y seguiremos usando este término para distinguirla. Y segundo, cuando A es la matriz de Lorentz y corresponde a la "polaridad" respecto al círculo unitario que apareció desde el capítulo 2. En la primera se tiene que un punto nunca está en su línea polar (de hecho es la línea más lejana a él de acuerdo con la métrica proyectiva), pero de la segunda podemos generalizar el principio fundamental.

Lema 9.1 *Si $\mathcal{C}$ es no degenerada y no vacía, entonces la línea polar de un punto en $\mathcal{C}$ es su línea tangente.*

Demostración. Estamos suponiendo que $\det A \neq 0$ y además que tenemos $[\mathbf{u}] \in \mathcal{C}$, es decir, tal que $\mathbf{u} \bullet_A \mathbf{u} = 0$; pues $\mathcal{C}$ está determinada por la ecuación $\mathbf{x} \bullet_A \mathbf{x} = 0$.

Sea $\ell = [\mathbf{u}]^A$; queremos demostrar que $\ell \cap \mathcal{C} = [\mathbf{u}]$. Supongamos que no; que existe $[\mathbf{v}] \neq [\mathbf{u}]$ tal que $[\mathbf{v}] \in \ell \cap \mathcal{C}$. Vamos a probar que entonces $\ell \subset \mathcal{C}$. Los puntos de ℓ se escriben como $[\alpha\,\mathbf{u} + \beta\,\mathbf{v}]$, y para cualquier $\alpha, \beta \in \mathbb{R}$ se tiene, por la simetría y bilinearidad, que

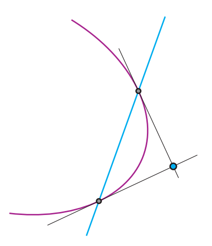
$$(\alpha\,\mathbf{u} + \beta\,\mathbf{v}) \bullet_A (\alpha\,\mathbf{u} + \beta\,\mathbf{v}) = \alpha^2\,(\mathbf{u} \bullet_A \mathbf{u}) + 2\alpha\beta\;(\mathbf{u} \bullet_A \mathbf{v}) + \beta^2\,(\mathbf{v} \bullet_A \mathbf{v})\,.$$

El primer sumando es 0 pues $[\mathbf{u}] \in \mathcal{C}$, el segundo también lo es pues $[\mathbf{v}] \in \ell \Leftrightarrow \mathbf{u} \bullet_A \mathbf{v} = 0$, y el tercero es cero al suponer que $[\mathbf{v}] \in \mathcal{C}$. Por tanto $\ell \subset \mathcal{C}$, y como contiene una recta entonces $\mathcal{C}$ es degenerada. La contradicción implica el lema. □

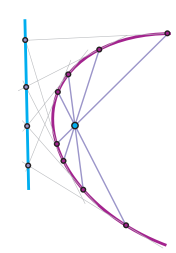
Con base en este lema y en la simetría polar (9.2) es fácil extraer las conclusiones análogas al caso del círculo. Suponiendo que A define una cónica no vacía (en $\mathbb{R}^2$ puede ser elipse, parábola o hipérbola, y en $\mathbb{P}^2$ simplemente no vacía y no degenerada), se cumple entonces que:

1. Los puntos polares de líneas concurrentes son colineales, e inversamente, las líneas polares de puntos colineales son concurrentes.

2. El punto polar de una secante a $\mathcal{C}$ se obtiene intersectando las tangentes a los dos puntos donde corta a $\mathcal{C}$; y podemos decir que este punto está *afuera* de $\mathcal{C}$. E inversamente, si $[\mathbf{u}]$ está afuera de la cónica entonces su recta A-polar $[\mathbf{u}]^A$ la intersecta en dos puntos (llamados los *pies* de $[\mathbf{u}]$ en la cónica $\mathcal{C}$) cuyas tangentes pasan por $[\mathbf{u}]$; así que para encontrar analíticamente los pies de un punto se encuentra la ecuación de su A-polar y se resuelve la ecuación cuadrática que da sus intersecciones con la cónica.



3. Si $[\mathbf{u}]$ está *dentro* de (es decir, ni afuera ni en) la cónica, entonces su recta A-polar $[\mathbf{u}]^A$ no la intersecta, está afuera y es la recta por los puntos polares a dos líneas cualquiera que pasan por $[\mathbf{u}]$. En particular, las intersecciones de las tangentes a los puntos extremos de cuerdas concurrentes son colineales (independientemente de dónde esté el punto de concurrencia).



EJERCICIO 9.1 Encuentra los pies de .... a la cónica ....

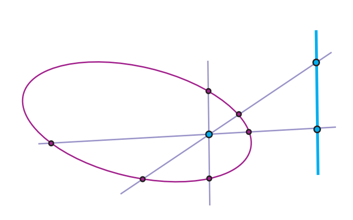
EJERCICIO 9.2 Demuestra que el centro de una cónica, cuando lo tiene, es un invariante afín.

9.3 Armonía y el grupo de invariancia

Toda cónica $\mathcal{C}$ (entenderemos por esto que es no vacía y no degenerada) tiene un grupo de transformaciones proyectivas que la deja invariante. Éste es conjugado del grupo

hiperbólico bajo una transformación proyectiva $\mathsf{g} \in \Pr(2)$ que manda a $\mathcal{C}$ en el círculo unitario (g existe por el Teorema de clasificación de cónicas proyectivas del capítulo 7). Así que se puede considerar al interior de la cónica $\mathcal{C}$ como el plano hiperbólico con su grupo de invariancia proyectiva como grupo de isometrías. Nótese que, puesto que g preserva tangencia, también preserva polaridad, y entonces podemos reinterpretar la polaridad en términos clásicos de armonía.

Puesto que la línea polar del centro, $\mathbf{o}$ digamos, del círculo unitario es la recta al infinito, se cumple que toda recta ℓ por $\mathbf{o}$ intersecta a su recta polar $\mathbf{o}^{\perp}$ en un par armónico respecto a sus intersecciones con el círculo (en una línea afín, el punto medio de un segmento es conjugado armónico del punto al infinito respecto a los extremos); es decir, $\mathbf{o} \in \ell$ y $\ell \cap \mathbf{o}^{\perp}$ es un par armónico respecto a la pareja $\ell \cap \mathbb{S}^1$. Como se puede mandar proyectivamente a la terna $\mathbb{S}^1, \mathbf{o}, \ell$ en cualquier terna que consiste en una cónica $\mathcal{C}$, un punto c en su interior y una recta por él; y las transformaciones proyectivas preservan armonía (Teorema 6.6), se tiene la siguiente:



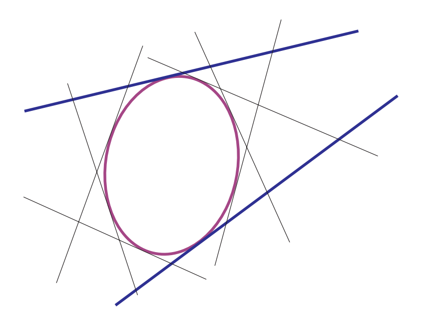
Proposicion 9.2 *Sea c un punto interior de una cónica (no vacía y no degenerada) $\mathcal{C}$. Entonces la recta polar de c respecto a $\mathcal{C}$ está formada por los conjugados armónicos de c respecto a la pareja $\ell \cap \mathcal{C}$, donde ℓ corre por las rectas que pasan por c; es decir, si $\mathsf{c} \in \ell$ entonces $\mathsf{c}, \ell \cap \mathsf{c}^{\perp}$ es armónico del par $\ell \cap \mathcal{C}$.*

Una manera, quizá más bonita, de entender esto es observar que el análogo de una cónica (no vacía y no degenerada) en la línea proyectiva, es decir, en dimensión 1, es una pareja de puntos. Respecto a ellos hay polaridad en $\mathbb{P}^1$ que es precisamente la relación de armonía; a cada punto se le asocia su conjugado armónico. Lo previo equivale a decir que al restringir la polaridad respecto a una cónica en $\mathbb{P}^2$ a sus líneas proyectivas se obtiene la polaridad ahí respecto a la intersección. Polaridad y armonía son instancias de lo mismo pero en diferentes dimensiones. Y debe ser claro que al subir la dimensión, esto se exiende. Para superficies cuadráticas en $\mathbb{P}^3$, los puntos tienen planos polares formados por los puntos en las líneas por él que son sus armónicos con respecto a la intersección con la msuperficie; las líneas tienen líneas polares o armónicas y los planos tienen puntos polares.

EJERCICIO 9.3 Define algebraicamente el *plano polar* de un punto respecto a una superficie cuadrática en $\mathbb{R}^3$.

9.4 Cónicas y proyectividades de rectas

Otro resultado del capítulo anterior que se generaliza fácilmente es el de la correspondencia tangencial. Dada una cónica $\mathcal{C}$ (no vacía y no degenerada) se tiene que todas sus líneas tangentes se identifican natural y canónicamente. Si ℓ_1 y ℓ_2 son dos tangentes a $\mathcal{C}$, su correspondenica tangencial es la transformación entre ellas que completa continuamente a:

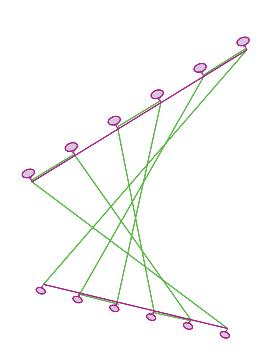


$$\begin{array}{rcl} \ell_1 & \leftrightarrow & \ell_2, \\ \ell \cap \ell_1 & \leftrightarrow & \ell \cap \ell_2, \end{array}$$

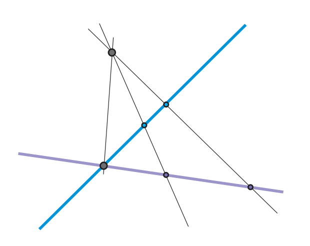
donde ℓ corre por el resto de las tangentes de $\mathcal{C}$. Que esta transformación es proyectiva se sigue de lo que ya hicimos (Proposición 8.8) al identificar proyectivamente a $\mathcal{C}$ con el círculo unitario. Pero surge una pregunta interesante que es la inversa, y cuya respuesta nos dará la oportunidad de colar de refilón resultados clásico que no habían tenido cabida en el texto.

Supongamos que ℓ_1 y ℓ_2 son dos rectas en $\mathbb{P}^2$ y que $f : \ell_1 \to \ell_2$ es una transformación proyectiva entre ellas. ¿Será cierto que las líneas $\langle p, f(p) \rangle$ con $p \in \ell_1$ son las tangentes a una cónica? Es decir ¿es toda transformación proyectiva entre rectas una correspondencia tangencial? Por supuesto que debemos eliminar el caso con el que empezamos a estudiar la geometría proyectiva que es la proyección desde un punto. Nótese que al proyectar desde un punto, la intersección $p_0 := \ell_1 \cap \ell_2$ se queda en su lugar, así que si añadimos la hipótesis de que $f(p_0) \neq p_0$ (que es el único punto donde podría pasar que $\langle p, f(p) \rangle$ no es una línea), la pregunta adquiere la dimensión de "teorema a la mano".

Antes de enunciarlo, recordemos la evidencia del "trazo artesanal de paraboloides" que algunos tuvimos la suerte de conocer desde la escuela secundaria. Si se toman dos intervalos no paralelos en una madera, se clavan clavitos a intervalos regulares en cada uno de ellos y luego se va entrelazando un hilo que una tachuelas correspondientes y de tal manera que se vaya cruzando (en un intervalo va en una dirección y en el otro al revés), se delinea una parábola, o, con un poco de imaginación tridimensional, parece un reglado de un paraboloide hiperbólico. Éste es un caso particular del problema anterior, la transformación entre las dos líneas es afín, y al corresponderse sus puntos al infinito, la línea por ellos es tangente a la cónica y por tanto es una parábola (dando ya por un hecho que la respuesta es afirmativa).



Teorema 9.3 *Sean ℓ_1 y ℓ_2 dos líneas en $\mathbb{P}^2$ que se intersectan en p_0, y sea $f : \ell_1 \to \ell_2$ una transformación proyectiva entre ellas. Si $f(p_0) \neq p_0$, entonces f es la correspondencia tangencial de una cónica; y si $f(p_0) = p_0$, f es la proyección desde un punto.*



Supongamos primero que $f(p_0) = p_0$. Consideremos otros dos puntos $a, b \in \ell_1$, y sea $q = \langle a, f(a)\rangle \cap \langle b, f(b)\rangle$. La proyección desde q de ℓ_1 a ℓ_2 coincide con f en tres puntos y por lo tanto (Teorema de 3 en 3) es f.

Supongamos entonces que $f(p_0) \neq p_0$. Notemos primero que la cónica que estamos buscando (la que dará la correpondencia tangencial entre ℓ_1 y ℓ_2) tiene como tangentes a ℓ_1 y a ℓ_2. Consideremos tres nuevas rectas ℓ_3, ℓ_4 y ℓ_5 de la forma $\langle p, f(p)\rangle$ con $p \in \ell_1$; si encontramos una cónica $\mathcal{C}$ que tenga a estas cinco rectas como tangentes habremos terminado: la correspondencia tangencial de ℓ_1 a ℓ_2 respecto a $\mathcal{C}$ coincide con f pues lo hace en tres puntos. Obsérvese entonces que estamos demostrando que el conjunto de rectas $\{\langle p, f(p)\rangle \mid p \in \ell_1\}$ son las tangentes a una cónica.

Queda por demostrar que hay una cónica tangente a cinco rectas dadas. Lo haremos dualizando un teorema clásico.

Teorema 9.4 *Por cinco puntos (distintos) en $\mathbb{P}^2$ pasa una curva cuadrática; ésta es única si y sólo si no hay cuatro de ellos colineales, y es no degenerada (una cónica) si y sólo si están en posición general.*

Demostración. Si no están en posición general, tres de los puntos son colineales y la recta ℓ que pasa por ellos tiene que estar en la cónica (de no ser así, la intersectaría en tres puntos). Además, ℓ junto con la recta que pasa por los otros dos es la cónica deseada (y degenerada); y está claro que no es única sólo cuando un cuarto, o todos, están en ℓ, pues entonces la segunda recta se puede hacer variar. Preocupémonos entonces por la existencia en general.

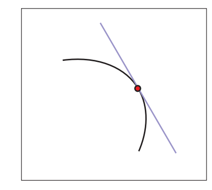
Podemos suponer que los cinco puntos están en $\mathbb{R}^2$. Consideremos las seis constantes del polinomio general de segundo grado como incógnitas. Al evaluar con las coordenadas de uno de los puntos dados e igualar a 0 (queremos que el punto anule el polinomio) se obtiene una ecuación lineal homogénea en las incógnitas. Tenemos entonces cinco ecuaciones lineales homogéneas con seis variables. Tienen una solución no trivial: esos valores nos dan un polinomio que se anula en los cinco puntos dados y define la cónica que buscamos.

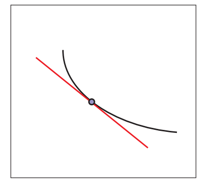
Formalmente hay un problemita; en este libro nunca llegamos a sistemas lineales de 5 por 6. Si se nos cree esta última afirmación, o ya se manejan rudimentos de álgebra lineal, que ahí quede el asunto; pero por puro prurito matemático una demostración que sólo usa elementos intrínsecos se describe en los ejercicios.

La unicidad equivale a que las soluciones tengan dimensión uno, pues dan entonces múltiplos de un polinomio y definen la misma cónica. Es más latoso de ver que esto equivale al enunciado del teorema. Los ejercicios de la demostración alternativa realzan entonces su valía. □

Regresando al Teorema 9.3, las cinco rectas corresponden a cinco puntos en el plano proyectivo dual, $\left(\mathbb{P}^2\right)^*$, que conviene pensar por separado de $\mathbb{P}^2$ aunque acabemos usando tres coordenadas proyectivas en cada uno de ellos. (Aquí estamos usando la dualidad clásica, donde a una recta le corresponde el punto $[\mathbf{v}]$ tal que la ecuación $\mathbf{v}\cdot\mathbf{x}=0$ la define, y que hemos denotado $\ell_{[\mathbf{v}]}\longleftrightarrow[\mathbf{v}]$.) Además, estos cinco puntos están en posición general, pues por la construcción es fácil ver que no hay tres de las rectas concurrentes (lo contrario implicaría que $f(p_0)=p_0$). Por el Teorema 9.4 existe una cónica $\mathcal{C}'$ no degenerada, en $\left(\mathbb{P}^2\right)^*$, que pasa por los cinco puntos. Tiene una matriz simétrica asociada A con $\det A\neq 0$. Para concluir, vamos a demostrar que la cónica asociada a la matriz A^{-1} es tangente a las cinco rectas originales.

Supongamos que $[\mathbf{v}]\in\left(\mathbb{P}^2\right)^*$ está en la cónica $\mathcal{C}'$, es decir que $\mathbf{v}^\top A\mathbf{v}=0$. Sabemos que la recta tangente a $\mathcal{C}'$ en $[\mathbf{v}]$ está dada por la ecuación $(A\mathbf{v})\cdot\mathbf{x}=0$. Pero además corresponde a un punto en $\mathbb{P}^2$; precisamente a $[A\mathbf{v}]\in\mathbb{P}^2$, que por la dualidad es un punto en la recta $\ell_{[\mathbf{v}]}$ (piénsese en una de las cinco originales). Tenemos entonces que las rectas tangentes a $\mathcal{C}'$ nos definen puntos en las rectas originales (los duales de sus tangentes), y estos puntos cumplen





$$(A\mathbf{v})^\top A^{-1}(A\mathbf{v})=\mathbf{v}^\top\left(A^\top A^{-1}A\right)\mathbf{v}=\mathbf{v}^\top A^\top\mathbf{v}=\mathbf{v}^\top A\mathbf{v}=0.$$

Entonces, si $\mathcal{C}$ es la cónica asociada a la matriz A^{-1}, su tangente en el punto $[A\mathbf{v}]$ es $\ell_{[A^{-1}(A\mathbf{v})]}=\ell_{[\mathbf{v}]}$. Y queda demostrado el Teorema 9.3.

La generalidad de estos últimos párrafos es para poder concluir con algo que era de esperarse.

Teorema 9.5 *Las tangentes a una cónica (no vacía y no degenerada) forman una cónica en el plano proyectivo dual.*

EJERCICIO 9.4 Demuestra, resolviendo el sistema de ecuaciones 5×6 correspondiente, que una cónica $\mathcal{C}$ pasa por los puntos $(0,0),(1,0),(0,1),(1,1)$, si y sólo si está definida por un polinomio de la familia

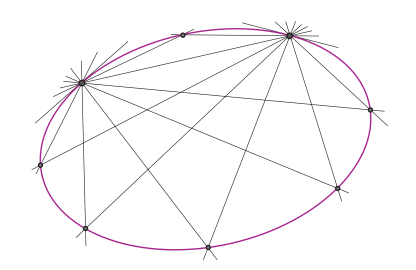
$$a\,x\,(x-1)+c\,y\,(y-1)$$

con $[a:c]\in\mathbb{P}^1$.

EJERCICIO 9.5 Demuestra que cualquier $p\in\mathbb{P}^2$ distinto de los cuatro puntos anteriores está en exactamente una de las cónicas de la familia. Concluye el Teorema 9.4 usando el de 4 en 4.

9.5 El Teorema místico de Pascal.

El dual del Teorema 9.3 da que si tenemos una transformación proyectiva entre dos haces de rectas (concurrentes), entonces los puntos de intersección de los pares de rectas correspondientes forman una cónica o una recta. Esto implica que las cónicas pueden definirse axiomáticamente sin uso de coordenadas o ecuaciones, son algo natural dentro de la geometría proyectiva, pues solo dependen de la relacion de incidencia. Veamos. Las proyecciones desde un punto entre rectas (con lo que empezamos a estudiar geometría proyectiva) se definen por incidencia (rectas que pasan por dos puntos e intersecciones de rectas), a veces se les llama *perspectividades*. Sin embargo, la composición de perspectividades no necesariamente es una perspectividad, es una transformación proyectiva o *proyectividad*. Estas nociones tienen sentido en el dual. Entonces, si tomamos los puntos $\ell \cap f(\ell)$ al correr ℓ en un haz concurrente (una línea dual) y siendo f una proyectividad con otro haz, se obtiene, en general, una cónica. Esta sería su definición sintética. Hacemos enfásis en esto al final para recordar y dar evidencia de lo que decíamos al principio: que también se puede hacer geometría a partir de axiomas y se llega a lo mismo;[1] la "geometría" es una.

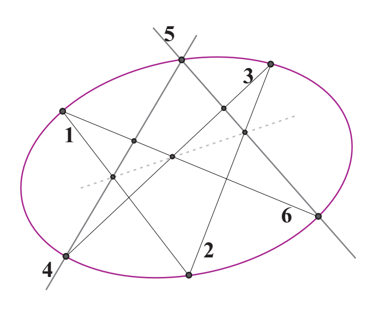


Para concluir, demostramos un teorema clásico usando el inverso de lo anterior, que las cónicas definen proyectividades entre los haces de rectas por sus puntos.

Teorema 9.6 (Pascal) *Los lados opuestos de un hexágono inscrito en una cónica se intersectan en puntos colineales.*

Demostración. Consideremos seis puntos $p_1, p_2, \ldots, p_6$ en una cónica $\mathcal{C}$. Nótese que el orden en que los dimos es independiente del orden cíclico en que aparecen en la cónica; el que usamos en la figura es simplemente para que los puntos relevantes queden cerca. Denotemos por ℓ_i a la línea que va del punto p_i al siguiente (según el hexágono), es decir, $\ell_i := \langle p_i, p_{i+1} \rangle$, donde $p_7 = p_1$; nótese que las seis líneas son distintas pues estamos suponiendo (con nuestra acepción de la palabra *cónica*) que $\mathcal{C}$ es no degenerada. Y denotemos $p_{ij} := \ell_i \cap \ell_j$, para $i \neq j$; de tal forma que $p_i = p_{(i-1)i}$. Debemos demostrar que los puntos p_{14}, p_{25}, p_{36} son colineales.

Sea $f : \ell_4 \to \ell_5$ la transformación proyectiva que se obtiene identificando a ℓ_4 con el haz concurrente en p_1 ($\ell_4 \ni p \mapsto \langle p, p_1 \rangle$), luego a éste con el haz de p_3 a través de la cónica $\mathcal{C}$ y finalmente a éste haz con ℓ_5 (por intersección). Nótese que f se puede pensar como proyectar ℓ_4 a $\mathcal{C}$ desde p_1 seguida de proyectar $\mathcal{C}$ a ℓ_5 desde p_3, y tenemos entonces los siguientes datos de f:



$$\begin{array}{ccccc} p_4 & \mapsto & p_4 & \mapsto & p_{35} \\ p_{14} & \mapsto & p_2 & \mapsto & p_{25} \\ p_{46} & \mapsto & p_6 & \mapsto & p_6 \\ p_5 & \mapsto & p_5 & \mapsto & p_5 \end{array}$$

[1] Este enfoque axiomático puede verse en [21].

Puesto que $f(p_5) = p_5 = \ell_4 \cap \ell_5$, se tiene que f es una proyección; y entonces es la proyección desde el punto

$$\langle p_4, p_{35} \rangle \cap \langle p_{46}, p_6 \rangle = \ell_3 \cap \ell_6 = p_{36} \,.$$

Como $f(p_{14}) = p_{25}$, se tiene entonces que p_{14}, p_{25} y p_{36} son colineales. □

El Teorema de Pascal generaliza al de Papus (6.13), pues éste último puede pensarse como un caso especial para cónicas degeneradas. Aunque la demostración original de Pascal se desconoce, está enunciado en el *Ensayo sobre las cónicas* que publicó a los 17 años; es evidencía intelectual de que sí conocía el trabajo pionero sobre geometría proyectiva de Desargues; además, éste y su padre fueron amigos, así que por lo menos lo conoció bien. A veces se le añade el adjetivo "*místico*" al Teorema: quizá por lo del *"hexagrama místico"* o bien porque el mismo Pascal fué absorvido más adelante en su vida por el misticismo.

EJERCICIO 9.6 Demuestra el Teorema de Papus (6.13) con la idea que se usó para demostrar el de Pascal. ¿Puedes enunciar y demostrar un Teorema de Papus-Pascal? Sólo falta analizar unos casos degenerados.

Capítulo 10

Apéndice

Este apéndice consta de dos secciones independientes. En la primera, establecemos muy brevemente las nociones básicas, la terminología y la notación de teoría de conjuntos que se usa en el libro. En la segunda, de un nivel mucho más alto, introducimos los números complejos apenas lo suficiente para enunciar el Teorema Fundamental del Álgebra. No damos su demostración, pero indicamos a grandes rasgos como probar el corolario de él que se usa en el capítulo 7.

10.1 Conjuntos

La propiedad básica de un conjunto A es que, dado cualquier objeto x, se puede decir si x *es elemento de* A, escrito $x \in A$ y que se lee "x está en A", o bien, si no lo es, lo cual se escribe $x \notin A$.

Para definir un conjunto hay que hacer explícito de alguna manera cuáles son sus *miembros* o *elementos*. Esto se puede hacer, al menos, de tres maneras. Primero, mediante una lista explícita, por ejemplo, los números naturales son

$$\mathbb{N} := \{0, 1, 2, 3, \cdots\}.$$

Segundo, dando una propiedad o propiedades que cumplen sus elementos; y tercero, mediante un parámetro que recorre los elementos de otro conjunto previamente definido. Por ejemplo, los números naturales pares se pueden definir por esa propiedad (la de ser par) o bien mediante un parámetro que recorre a los naturales:

$$\{n \in \mathbb{N} \mid n \text{ es par}\} = \{2n \mid n \in \mathbb{N}\};$$

la barra de enmedio " | " se lee "tal que" o bien "donde"; en algunos textos se usan dos puntos " : " en vez de " | ".

La *unión* y la *intersesección* de dos conjuntos A y B son respectivamente:

$$A \cup B = \{x \mid x \in A \text{ o } x \in B\} \qquad \text{y} \qquad A \cap B = \{x \mid x \in A \text{ y } x \in B\}.$$

Cuando todos los elementos de un conjunto A también son elementos de B, se dice que A es *subconjunto* de B, o bien que B *contiene* a A, y se escribe $A \subset B$, o bien $B \supset A$. De tal manera que siempre se tiene que $A \cup B$ contiene a A y a B, y que $A \cap B$ es subconjunto de A y de B. Otra *operación* de conjuntos (cómo obtener un conjunto a partir de otros dos) es la *diferencia*, que se denota $A \backslash B$ o bien $A - B$, que es

$$A \backslash B = A - B := \{x \in A \mid x \notin B\}.$$

Un conjunto importantísimo es el *conjunto vacío,* que se denota $\emptyset$, y es el único conjunto que no tiene elementos; para cualquier x se tiene que $x \notin \emptyset$. Así que $A - B$ cumple que $B \cap (A - B) = \emptyset$ y $(A - B) \subset A$. Además, el vacio es subconjunto de cualquier otro conjunto, pues por "*vacuidad*" se cumple para cualquier conjunto A que

$$x \in \emptyset \quad \Rightarrow \quad x \in A\,,$$

donde "$\Rightarrow$" significa "*implica que*"; puesto que $x \in \emptyset$ nunca sucede, cada vez que sucede se cumple lo que sea.

La última operación de conjuntos que usaremos es el *producto cartesiano.* A partir de dos conjuntos A y B, se puede construir un nuevo conjunto $A \times B$, llamado su *producto* que consta de todas las parejas ordenadas con el primer elemento en A y el segundo en B, es decir

$$A \times B := \{(a, b) \mid a \in A,\ b \in B\}.$$

Relaciones

Una *relación* R en un conjunto A es un subconjunto de su producto cartesiano, $R \subset A \times A$; es decir, es distinguir una colección de parejas *relacionadas.* Pero es más comodo y usual denotarla con una tilde " $\sim$ " (o un símbolo parecido) con la convención

$$a \sim b \quad \Leftrightarrow \quad (a, b) \in R,$$

donde "$\Leftrightarrow$" significa "*si y sólo si*"; y cuando $a \sim b$ se dice, y se lee, que a "*está relacionada con*" b. Vale entonces la notación $a \nsim b$ para $(a, b) \notin R$: cuando a no está relacionado con b.

Hay dos tipos de relaciones que con frecuencia surgen naturalmente. Una relación en A es una *relación de orden* si cumple las siguientes propiedades:

i) $a \sim a$ para todo $a \in A$ (*reflexividad*);

ii) $a \sim b$ y $b \sim a \quad \Rightarrow \quad a = b$;

iii) $a \sim b$ y $b \sim c \quad \Rightarrow \quad a \sim c$ (*transitividad*).

Un ejemplo claro de orden son los números naturales con la relación "*ser menor o igual*" denotada " $\leq$ ". Otro es la relación de *contención* en el conjunto de los subconjuntos de un conjunto A; tienen entonces sentido las definiciones alternas "la unión de A y B es el mínimo conjunto que contiene a ambos" y "la intersección el máximo subconjunto común". Obsérvese, además, que una relación de orden siempre tiene una relación de *orden estricto* asociada donde se eliminan de la relación a las *igualdades* $a \sim a$ y se llama entonces *antireflexiva*. Para el orden en los números se obtiene "*ser menor*" denotada " $<$ "; y para la de contención se obtiene la de "*contención estricta*", $\subsetneq$.

Una relación $\sim$ en A es de *equivalencia* si es reflexiva, transitiva y además cumple:

iv) $a \sim b \quad \Rightarrow \quad b \sim a$ (*simetría*).

Cuando se tiene una relación de equivalencia en un conjunto A, los subconjuntos máximos tales que todos sus elementos están relacionados entre sí se llaman *clases de equivalencia*; y el conjunto A se *parte* en clases de equivalencia; es decir, cualquier elemento $a \in A$ está en una clase de equivalencia, la de los equivalentes a a, y las clases de equivalencia son disjuntas. E inversamente, una *partición* de un conjunto en subconjuntos ajenos y que lo cubren induce una relación de equivalencia en él.

El otro concepto que, por elemental, debería entrar en este apéndice es el de "función", pero es tan importante para el texto que decidimos incluirlo al principio del capítulo 3.

10.2 Números complejos

Los números complejos, denotados $\mathbb{C}$, se construyen a partir de los números reales $\mathbb{R}$ al introducir, o anexar, un nuevo número que se denota i, que cumple

$$i^2 = -1.$$

De tal manera que los números complejos se pueden definir como

$$\mathbb{C} := \{a + b\,i \mid a, b \in \mathbb{R}\}.$$

Por la igualdad $i^2 = -1$, la suma y la multiplicación de los reales se extiende naturalmente a $\mathbb{C}$, de tal manera que se convierte en un *campo* (ver sección 1.3.1).

Un polinomio en $\mathbb{C}$ es una expresión de la forma

$$P(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n, \tag{10.1}$$

con sus *coeficientes* $c_0, c_1, \ldots, c_n \in \mathbb{C}$, y es de *grado* n si $c_n \neq 0$. De tal manera que para cualquier $z \in \mathbb{C}$ se puede *evaluar* $P(z)$ (desarrollando la expresión) y de nuevo es un número complejo. Una *raíz* de P es un número complejo que anula al polinomio,

es decir, tal que al evaluar a P en él da 0. Por ejemplo, las raices del polinomio x^2+1 son i y $-i$.

El **Teorema Fundamental del Álgebra**, probado por Gauss, dice que un polinomio de grado n tiene n raíces, o más precisamente, que dado el polinomio (10.1) (con $c_n \neq 0$) existen $z_1, z_2, \ldots, z_n \in \mathbb{C}$ (sus raíces) tales que

$$P(x) = c_n (x - z_1)(x - z_2) \cdots (x - z_n).$$

Un campo con esta propiedad se llama *algebraicamente cerrado*. Los reales no lo son pues el polinomio x^2+1 no tiene raíces reales; pero al adjuntar una raíz de él se cierra.

El corolario de este teorema que se usa en el libro es que *un polinomio real de grado impar tiene al menos una raíz real*. Se sigue de la siguiente observación.

Para cada número complejo $z = a + b\,i$ se tiene su *conjugado*

$$\bar{z} = a - b\,i,$$

y es fácil ver que esta asignación cumple las siguientes propiedades:

$$\begin{aligned} \overline{(z+w)} &= \bar{z} + \bar{w} \\ \overline{(zw)} &= \bar{z}\,\bar{w} \\ z = \bar{z} &\Leftrightarrow z \in \mathbb{R}. \end{aligned}$$

De tal manera que para un polinomio $P(x)$ con todos sus coeficientes reales se tiene que $\overline{P(z)} = P(\bar{z})$ para cualquier $z \in \mathbb{C}$. En particular, si z_0 es una raiz, tambien lo es su conjugado $\bar{z}_0$, pues $P(z_0) = 0 = \overline{P(z_0)} = P(\bar{z}_0)$. Es decir, las raíces complejas (no reales) de un polinomio con coeficientes reales vienen en parejas conjugadas. Y por lo tanto, si tiene grado impar tiene al menos una raíz real.

Bibliography

[1] Silvio Bergia, *Einstein and the birth of special relativity.* En *Einstein: A Centenary Volume,* A. P. French, Editor. Harvard University Press, Cambridge, Mass. 1979.

[2] Javier Bracho, *¿En qué espacio vivimos?* La Ciencia para todos No. 77; Fondo de Cultura Económica, México D.F., 1989.

[3] John H. Conway y Derek A. Smith, *On quaternions and octonions.* A. K. Peters; Natick, Massachusets, 2003.

[4] H. S. M. Coxeter, *Regular polytopes.* Dover; Nueva York 1973.

[5] H. S. M. Coxeter, *Projective honeycombs. ?????*

[6] H. S. M. Coxeter, *Introduction to geometry,* Wiley, Nueva York 1961.

[7] H. S. M. Coxeter, *Projective geometry,* 2a ed., University of Toronto Press, 1974.

[8] Branko Grunbaum y G. C. Shepard, *Tilings and Patterns,* W. H. Freeman, Nueva York, 1987.

[9] David Hilbert y S. Cohn-Vossen, *Geometry and the imagination,* 1932. Traducción al ingles por P. Nemenyi, Chelsea Publishing Company; Nueva York 1952.

[10] Alan Holden, *Shapes, space and symmetry.* Dover; Nueva York 1971.

[11] L. Christine Kinsey y Teresa E. Moore, *Symmetry, shape and space.* Key College Publishing en coperación con Springer Verlag; Nueva York, 2002.

[12] Elmer Rees, *Elementary geometry, ???*

[13] R. von B. Rucker, *Geometry, relativity and the fourth dimension.* Dover, Nueva York, 1977.

[14] Patrick J. Ryan, *Euclidean and non-euclidean geometry; an analytic approach.* Cambridge University Press; Cambridge, 1986.

[15] Hermann Weyl, *Symmetry.* Princeton University Press; Princeton, 1952.

[16] Robert Goldblat, *Orthogonality and spacetime geometry.* Univesitext, Springer-Verlag, Nueva York, 1987.

[17] George A. Jennings, *Modern geometry with applications.* Univesitext, Springer-Verlag, Nueva York 1994.

[18] Benoit Mandelbrot, *Fractals everywhere*

[19] *Generating fractals*

[20] David Darling, *The universal book of mathematics*: *from abracadabra to Zeno's paradoxes.* John Wiley & Sons, Inc., New Jersey 2004.

[21] Judith N. Cederberg, *Acourse in modern geometries.* Undergraduate texts in Mathematics, Springer-Verlag, Nueva York, 1989.

[22] John Stillwell, *Mathematics and its history.* Second Edition. Springer-Verlag, Nueva York 2002, 1989.

[23] George E. Martin, *Transformation geometry; an introduction to symmetry.* Undergraduate texts in Mathematics, Springer-Verlag, Nueva York, 1982.

[24] David Mumford, Caroline Series y David Wright, *Indra's pearls; the vision of Felix Klein.* Cambridge University Press, Cambridge, 2002.

Index